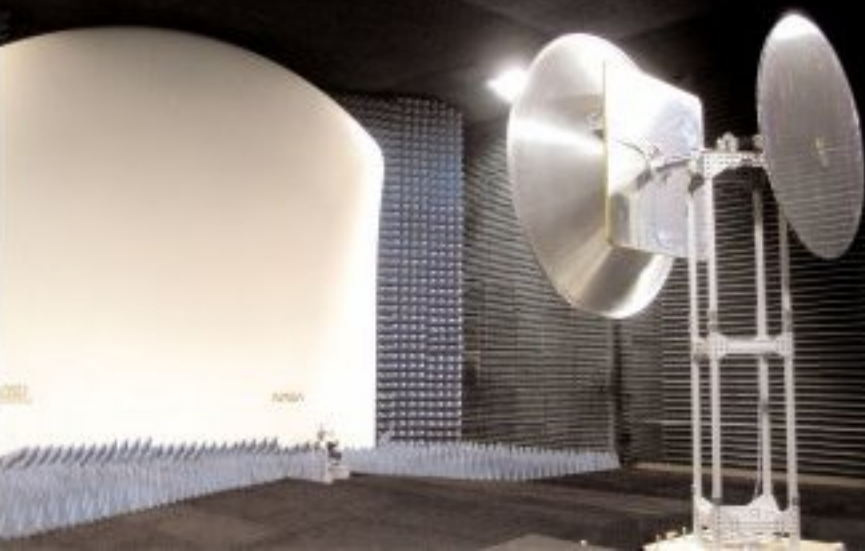


The background of the cover features a large, complex antenna array structure, possibly a radio telescope or a satellite ground station, with multiple large parabolic dishes and a dense network of support beams and cables. The image is in grayscale, giving it a technical and scientific appearance.

DR. JOHN L. VOLAKIS

ANTENNA ENGINEERING HANDBOOK

FOURTH EDITION



P · A · R · T · 1

Introduction and Fundamentals

Chapter 1

Fundamentals of Antennas, Arrays, and Mobile Communications

Thomas F. Eibert

Universität Stuttgart

John L. Volakis

The Ohio State University

CONTENTS

1.1	INTRODUCTION	1-4
1.2	HUYGENS' AND EQUIVALENCE PRINCIPLES	1-5
1.3	HERTZIAN AND FITZGERALD ELEMENTARY RADIATORS	1-7
1.4	FAR-FIELD ANTENNA PROPERTIES, POWER TRANSFER, AND RECIPROCITY	1-8
1.5	ANTENNAS AS ELECTROMAGNETIC CIRCUITS	1-11
1.6	POLARIZATION	1-14
1.7	DIRECTIVITY PATTERNS FROM CONTINUOUS LINE SOURCES	1-17
1.8	DIRECTIVITY PATTERNS FROM AREA SOURCE DISTRIBUTIONS	1-21
1.9	FUNDAMENTALS OF ANTENNA ARRAYS	1-27
1.10	BASIC CONCEPTS IN MOBILE COMMUNICATIONS	1-32

1.1 INTRODUCTION*

Antennas are key components of any wireless communication system.^{1,2} They are the devices that allow for the transfer of a signal (in a wired system) to waves that, in turn, propagate through space and can be received by another antenna. The receiving antenna is responsible for the reciprocal process, i.e., that of turning an electromagnetic wave into a signal or voltage at its terminals that can subsequently be processed by the receiver. The receiving and transmitting functionalities of the antenna structure itself are fully characterized by Maxwell's equations and are fairly well understood.³ The *dipole antenna* (a straight wire, fed at the center by a two-wire transmission line) was the first antenna ever used and is also one of the best understood.^{1,2} For effective reception and transmission, it must be approximately $\lambda/2$ long (λ = wavelength) at the frequency of operation (or multiples of this length). Thus, it must be fairly long (or high) when used at low frequencies ($\lambda = 1$ m at 300 MHz), and even at higher frequencies (UHF and greater), its protruding nature makes it quite undesirable. Further, its low gain (2.15 dB), lack of directionality, and extremely narrow bandwidth make it even less attractive. Not surprisingly, the Yagi-Uda antenna (typically seen on the roof of most houses for television reception) was considered a breakthrough in antenna technology when introduced in the early 1920s because of its much higher gain of 8–14 dB. Log-periodic wire antennas introduced in the late 1950s and 1960s and wire spirals allowed for both gain and bandwidth increases. On the other hand, even today high gain antennas rely on large reflectors (dish antennas) and waveguide arrays (used for many radar systems).

Until the late 1970s, antenna design was based primarily on practical approaches using off-the-shelf antennas such as various wire geometries (dipoles, Yagi-Uda, log-periodics, spirals), horns, reflectors, and slots/apertures as well as arrays of some of these. The antenna engineer could choose or modify one of them based on design requirements that characterize antennas, such as gain, input impedance, bandwidth, pattern beamwidth, and sidelobe levels (see References 4 and 5 for a description of these quantities). Antenna development required extensive testing and experimentation and was, therefore, funded primarily by governments. However, in recent years, dramatic growth in computing speed and development of effective computational techniques for realistic antenna geometries has allowed for low-cost virtual antenna design. Undoubtedly, the explosive growth of wireless communications and microwave sensors, microwave imaging needs, and radars has been the catalyst for introducing a multitude of new antenna designs over the past decade and an insatiable desire for using modern computational techniques for low-cost designs. Requirements for

* Heinrich R. Hertz was the first to demonstrate the generation of radio waves at UHF using a gap dipole in 1885–1886 at Karlsruhe University (Germany). Hertz was able to detect radio waves 20 m away using a high-voltage electrical spark discharge to excite the dipole gap. From recorded conversations, Hertz did not seem to understand the impact of his experiments, but instead did it as a validation of the newly developed Maxwell's equations. Within ten years, Tesla at the Franklin Institute in the U.S., Marconi in Bologna, Italy, Popov in Russia, and Bose in India, demonstrated *wireless telegraphy*. In 1892, Tesla delivered a widely distributed presentation at the IRE of London about "transmitting intelligence without wires," and in 1895, he transmitted signals detected 50 miles (80 km) away. Concurrently, in 1894 Bose used wireless signals to ring a bell in Calcutta, and Popov presented his radio receiver to the Russian Physical & Chemical Society on May 7, 1895. Marconi is certainly considered the key individual for his contributions to the commercialization of radio waves, and he received the Nobel prize for his work in 1909. Nevertheless, Marconi's widely advertised first radio wave transmission experiment was in 1895, and his British patent application in 1897 was preceded by that of Tesla. A culmination of Marconi's experiments was the December 12, 1901, trans-Atlantic radio wave transmission of the Morse code for the letter S. The success of this experiment is often disputed, possibly due to strong atmospheric noise during the time of the experiment, but by the 1920s the U.S. had hundreds of radio stations, and in 1922, the BBC began transmitting in England. Subsequent development of radio detectors, vacuum tubes, and the tiny transistor in 1947 played a critical role in the practical everyday use of radio waves for communication and wireless transmission of information.

conformal antennas (non-protruding) for airborne systems, increased bandwidth requirements, and multifunctionality have led to heavy exploitation of printed (patch) or other slot-type antennas⁴ and the use of powerful computational tools (commercial and noncommercial) for designing such antennas.

Needless to say, the commercial mobile communications industry has been the catalyst for the recent explosive growth in antenna design needs. Certainly, the past decade has seen an extensive use of antennas by the public for cellular, GPS, satellite, wireless LAN for computers (WiFi), Bluetooth technology, Radio Frequency ID (RFID) devices, WiMAX, and so on. However, future needs will be even greater when a multitude of antennas are integrated into say automobiles for all sorts of communication needs and into a variety of portable devices and sensors for monitoring and information gathering. Certainly, future RFID devices will most likely replace the bar codes on all products while concurrently allowing for instantaneous inventorying. For military applications, there is an increasing need for small and conformal multifunctional antennas that can satisfy a plethora of communications needs using as little space as possible.

In this first chapter of the handbook, we provide a summary of antenna fundamentals and introduce antenna parameters typically used for characterizing antenna properties often employed to evaluate the entire radio system. We start with the radiation of an ideal (Hertzian) or infinitesimal dipole and proceed to the resonant $\lambda/2$ dipole, antenna arrays, and mobile communication concepts.

1.2 HUYGENS' AND EQUIVALENCE PRINCIPLES

The electromagnetic behavior and thus the functioning of antennas is governed by Maxwell's equations,³ which must be solved for a particular antenna and a given excitation. Typically, exact solutions of Maxwell's equations are not available and thus numerical modeling is often used to compute approximate solutions for practical configurations. A formal simplification of electromagnetic antenna problems can be achieved by employing the equivalence principle.³ If interest is restricted to the field solution in a limited region of space, the antenna configuration can be replaced by the equivalent electromagnetic sources located on the surface of a volume enclosing the antenna configuration (see Figure 1-1). Because the antenna materials are no longer there, these sources are usually radiating in a homogeneous solution space (such as free-space), and the corresponding fields can thus be calculated by evaluating the radiation integrals.

The equivalent sources are not uniquely defined, and there are many different ways of constructing them. In general, the equivalent sources are a composition of electric and magnetic surface current densities representing the excitation terms in Maxwell's equations. A straightforward way of constructing equivalent sources is provided by Huygens' principle.³ Huygens' principle states that the field solution in a region V is completely determined by the tangential fields over the surface S enclosing V . The corresponding electric and magnetic equivalent surface current densities are given by

Electric current density:

$$\mathbf{J} = \hat{n} \times \mathbf{H} \quad (1-1)$$

Magnetic current density:

$$\mathbf{M} = -\hat{n} \times \mathbf{E} \quad (1-2)$$

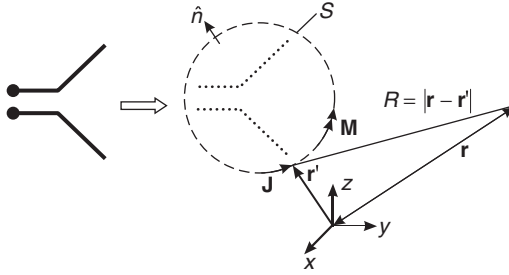


FIGURE 1-1 Replacement of an antenna by equivalent electric and magnetic surface current densities

where both \mathbf{J} and \mathbf{H} are expressed in amperes per meter (A/m) and \mathbf{M} and \mathbf{E} are expressed in volts per meter (V/m). For the problem of a radiating antenna, as illustrated in Figure 1-1, the outer boundary of V is assumed to be located at infinity, where the fields radiated by the corresponding equivalent sources can be neglected. As shown in the figure, the antenna can be replaced by equivalent sources on an arbitrary surface S enclosing it. As already mentioned, these equivalent sources reproduce the radiated fields of the antenna, and they can be assumed as radiating in homogeneous space. For a particular antenna configuration, the exact determination of \mathbf{J} and \mathbf{M} requires knowledge of the true field distribution on S . However, for many practical antennas, an approximate determination of \mathbf{J} and \mathbf{M} is possible. For instance, placing S to coincide with a metallic section of the antenna structure causes \mathbf{M} to vanish on these portions of S .

The radiated fields from any antenna can be obtained by integrating the field contributions of the equivalent electric and magnetic current densities using the well-known radiation integral:³

$$\mathbf{E} = -j\omega\mu_0 \oint_S \left[\mathbf{J}(\mathbf{r}') \frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} + \frac{1}{k_0^2} (\nabla' \cdot \mathbf{J}(\mathbf{r}')) \nabla \frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \right] ds' + \oint_S \mathbf{M}(\mathbf{r}') \times \nabla \frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} ds'$$

which for the far-field ($r \rightarrow \infty$) reduces to (see Figure 1-1)

$$\mathbf{E} = -j\omega\mu_0 \frac{e^{-jk_0r}}{4\pi r} \oint_S \left[(\bar{\mathbf{I}} - \hat{\mathbf{r}}\hat{\mathbf{r}}) \cdot \mathbf{J}(\mathbf{r}') - \sqrt{\frac{\epsilon_0}{\mu_0}} \hat{\mathbf{r}} \times \mathbf{M}(\mathbf{r}') \right] e^{jk_0\hat{\mathbf{r}} \cdot \mathbf{r}'} ds'$$

where

$\bar{\mathbf{I}}$ = unit dyad

\mathbf{r} = defines location of observation point (see Figure 1-1)

r = distance (in m) to observation point

\mathbf{r}' = defines location of the integrated surface current densities

$\hat{\mathbf{r}}$ = unit vector in radial direction

ϵ_0 = free-space permittivity

μ_0 = free-space permeability

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = \text{free-space impedance}$$

$$k_0 Z_0 = \omega \mu_0$$

$$k_0 = \beta = 2\pi / \lambda$$

$$\lambda = \text{wavelength (in meters, m)}$$

$$j = \sqrt{-1}$$

$$E \text{ is given in volts per meter (V/m)}$$

$$H \text{ is given in amperes per meter (A/m)}$$

For the ideal (delta) or infinitesimal electric (Hertzian) or magnetic (Fitzgerald) dipole sources, the radiation integrals are eliminated and the fields can be given in closed forms. The resulting field expressions can then be used to extract and study the usual antenna parameters.

1.3 HERTZIAN AND FITZGERALD ELEMENTARY RADIATORS

Considering the infinitesimal electric dipole $\mathbf{J} = \hat{z} Idz \delta(z)$, as illustrated in Figure 1-2, the resulting rms (root mean square) electric and magnetic field components are given by

$$\begin{aligned} E_r &= k_0^2 \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{Idz}{2\pi} \left[\frac{1}{(k_0 r)^2} - \frac{j}{(k_0 r)^3} \right] \cos\theta e^{-jk_0 r}, \\ E_\theta &= jk_0^2 \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{Idz}{4\pi} \left[\frac{1}{k_0 r} - \frac{j}{(k_0 r)^2} - \frac{1}{(k_0 r)^3} \right] \sin\theta e^{-jk_0 r}, \\ H_\phi &= jk_0^2 \frac{Idz}{4\pi} \left[\frac{1}{k_0 r} - \frac{j}{(k_0 r)^2} \right] \sin\theta e^{jk_0 r}, \\ F &= H = H = 0 \end{aligned} \quad (1-3)$$

where

Idz = moment of the differential current element
(I is given in rms amperes, and dz is given in meters)

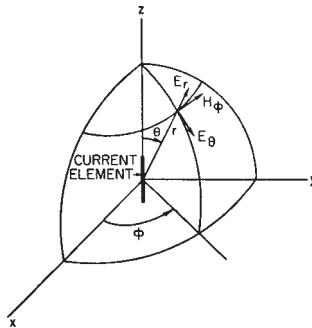


FIGURE 1-2 Coordinate system for an electric dipole

A time factor of $e^{j\omega t}$ has been suppressed since a sinusoidally time-varying current excitation of constant frequency is assumed. These are the exact fields, but antenna parameter evaluation is usually carried out using simplified far-fields, i.e., when r is much greater than the wavelength λ . Under these conditions, terms of order $1/r^2$ and greater are neglected to obtain

$$\begin{aligned} E_\theta &= jk_0 \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{Idz}{4\pi} \sin\theta \frac{e^{-jk_0 r}}{r} = \sqrt{\frac{\mu_0}{\epsilon_0}} H_\phi, \\ H_\phi &= jk_0 \frac{Idz}{4\pi} \sin\theta \frac{e^{-jk_0 r}}{r} = \sqrt{\frac{\epsilon_0}{\mu_0}} E_\theta. \end{aligned} \quad (1-4)$$

The complex Poynting vector \mathbf{S}^3 in the far-field is given by

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}^* = k_0^2 \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{Idz}{16\pi^2} \frac{\sin^2\theta}{r^2} \hat{r} \quad (1-5)$$

showing that it is a purely real quantity and indicating that power transport is in the r direction (away from the elementary current) without any reactive energy. Also, it is seen that the radiated power density (power flow per unit area) for any $r = \text{const.}$ depends on $\sin^2\theta$ (independent of ϕ). This is referred to as the radiation pattern plotted in dB. For an elementary (ideal or infinitesimal) magnetic dipole $\mathbf{M} = \hat{z} I_m dz \delta(z)$, the radiated fields can be found by duality³ (replacing E by H , H by $-E$, and I by I_m).

1.4 FAR-FIELD ANTENNA PROPERTIES, POWER TRANSFER, AND RECIPROCITY

Because antenna radiation can be represented by radiation integrals over equivalent current sources, as considered in the previous paragraph, no reactive field components will be found in the far-field of any antenna. Further, far-field antenna characterization can be performed by considering power flow under the constraint of energy conservation. However, the distance from an antenna to its far-field depends on the antenna, and it is commonly accepted that the far-field region starts after the distance

$$R = r = \frac{2D^2}{\lambda} \quad (1-6)$$

where D is the largest dimension of the antenna. This is due to the varying propagation distances of field contributions from different parts of the antenna to an observation point P , as illustrated in Figure 1-3. In the far-field, every antenna is considered a point source, and the far-field criterion in Eq. 1-6 is derived under the assumption that the phase errors due to the varying propagation distances are less than $\pi/8$.

Consider an antenna located at the origin of a spherical coordinate system, as illustrated for the electric current element in Figure 1-2. Assume that the *antenna is transmitting* and let

- P_t = power accepted by the antenna (in Watts)

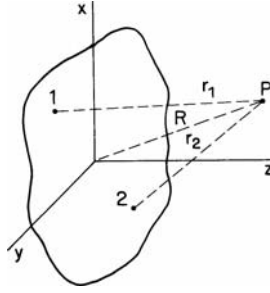


FIGURE 1-3 Schematic representation of an antenna aperture showing the observation point P and the distances to the observation point from two points on the antenna

- P_{rad} = power radiated by the antenna (in Watts)
- η = radiation efficiency (unitless)

These quantities are related by

$$\eta = \frac{P_{\text{rad}}}{P_t} \quad (1-7)$$

Let

- $S_i(\theta, \phi)$ = power density (in Watts/square meter, W/m^2)

and remark that $S_i(\theta, \phi)$ is independent of the distance from the antenna r , a characteristic of the far-field region. The total radiated power can be obtained by integrating the power density over a surface enclosing the antenna. Such a surface can be of any shape and can be as close to the antenna as desired. However, for simplicity, it is convenient to choose the surface to be a sphere, giving

$$P_{\text{rad}} = \int_0^{2\pi} \int_0^\pi S_i(\theta, \phi) r^2 \sin\theta \, d\theta \, d\phi \quad (1-8)$$

with the average power density being

$$P_{\text{avg}} = \frac{P_{\text{rad}}}{4\pi r^2} \quad (1-9)$$

Now, let

- $D_i(\theta, \phi)$ = directivity (unitless)

Directivity is a measure of the antenna to concentrate radiated power in a particular direction, and it is related to power density as

$$D_i(\theta, \phi) = \frac{S_i(\theta, \phi)}{P_{\text{avg}}} = \frac{S_i(\theta, \phi)}{P_{\text{rad}} / (4\pi r^2)} \quad (1-10)$$

The directivity of an antenna is the ratio of the achieved power density in a particular direction to that of an isotropic antenna.

Let

$$G_i(\theta, \phi) = \text{gain (unitless)}$$

The gain of an antenna is related to the directivity and radiated power density via the relation

$$G_i(\theta, \phi) = \eta D_i(\theta, \phi) = \frac{\eta S_i(\theta, \phi)}{P_{\text{rad}} / (4\pi r^2)} \quad (1-11)$$

and from Eq. 1-7

$$G_i(\theta, \phi) = \frac{S_i(\theta, \phi)}{P_i / (4\pi r^2)} \quad (1-12)$$

If you have a lossless antenna (i.e., $\eta = 1$), the directivity and the gain will be identical.

Now consider a *receiving antenna* exposed to a power density radiated from some transmitting antenna. The ability of the antenna to receive energy is quantified through

- $A_{e,r}(\theta, \phi)$ = effective area (in square meters)

where the antenna's location is assumed to be at the origin of the coordinate system.

Under the assumption of reciprocity,⁵ the effective area of an antenna is related to the gain via the relation

$$A_{e,r}(\theta, \phi) = \frac{\lambda^2}{4\pi} G_r(\theta, \phi) \quad (1-13)$$

where λ = wavelength. Note that reciprocity holds only for lossless antennas. Also, Eq. 1-13 depends on wavelength and therefore on frequency. Under these circumstances, antenna characterization can be performed either for the transmitting or receiving case with behavior for the other being immediately known.

Suppressing the angular dependencies of the transmitting and receiving antennas in their local coordinate systems, the received power is equal to the product of the power density of the incident wave and the effective aperture of the receiving antenna. That is

$$P_r = S_i A_{e,r}$$

Substituting from Eq. 1-12 and Eq. 1-13 yields

$$P_r = \frac{G_i P_i}{4\pi r^2} \frac{\lambda^2 G_r}{4\pi}$$

or

$$P_r = \left(\frac{\lambda}{4\pi r} \right)^2 G_i G_r P_i \quad (1-14)$$

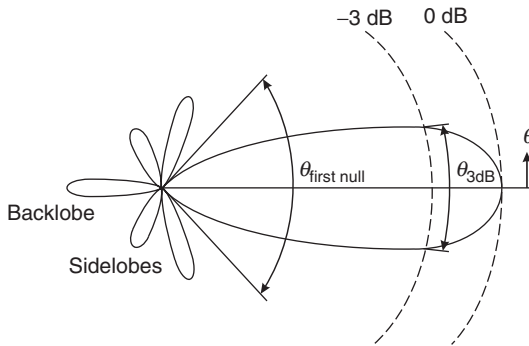


FIGURE 1-4 Antenna pattern in plane $\phi = \text{const}$

Note that G_t is the gain of the transmitting antenna in the direction of the receiving one and G_r is the gain of the receiving antenna in the direction of the transmitting one. A form of this equation was presented first by Friis⁶ and is usually called the *Friis transmission formula*.

The angular dependence of the radiating and receiving properties of an antenna in the far-field is often referred to as the *antenna radiation pattern*. Thus, a pattern is a normalized plot of the directivity, gain, or effective aperture as a function of angle and is often given in dB scale. Typically, the radiated normalized radiated power density or radiated field is plotted in dB (for the infinitesimal or ideal dipole, the power density $\sin^2 \theta$ is plotted in dB). A typical antenna pattern has a main lobe, sidelobes, minor lobes, a backlobe, and several nulls, as illustrated in Figure 1-4, in a $\phi = \text{const}$. plane. The half-power or 3 dB beamwidth of the main lobe (or main beam) is indicated in the drawing. If the pattern of an antenna is given in a plane parallel to the **E** field vector, the corresponding pattern is referred to as an **E** plane pattern. Alternatively, if the pattern cut is in a plane parallel to the **H** field polarization, it is called an **H** plane pattern.

There are many types of antenna radiation patterns, but the most common are

- Omnidirectional (azimuthal-plane)
- Pencil beam
- Fan beam
- Shaped beam

The *omnidirectional* pattern is most popular in communication and broadcast applications. The azimuthal pattern is circular, but the elevation pattern has some directivity to increase the gain in the horizontal direction. The term *pencil beam* is applied to a highly directive antenna pattern consisting of a major lobe contained within a cone of a small solid angle. Highly directive antenna patterns can be employed for point-to-point communication links and help reduce the required transmitter power. A *fan beam* is narrow in one direction and wide in the other. A *fan beam* is typically used in search or surveillance radars. *Shaped beam* patterns are adapted to the requirements of particular applications.

1.5 ANTENNAS AS ELECTROMAGNETIC CIRCUITS

A symbolic transition between a waveguide (transmission line) and an antenna is shown in Figure 1-5. In the case of a radiating antenna, a guided wave with amplitude a is traveling

toward the antenna and is more or less radiated into the surrounding medium. In the case of a receiving antenna, the received energy is transferred into a guided wave with amplitude b traveling down the waveguide away from the antenna. Consequently, from a circuit's viewpoint, the antenna can be considered as a one-port element provided only one guided mode exists at the port, as is normally the case; higher order modes can be considered as additional ports.

In the radiating mode, the one-port element is characterized by its complex reflection coefficient,

$$\Gamma = \frac{b}{a} \quad (1-15)$$

In the receiving mode, the received signal is converted into a traveling wave of amplitude b_0 . In general, the guided wave amplitude b at the antenna port is given by

$$b = \Gamma a + b_0 \quad (1-16)$$

If the antenna is a two-terminal circuit, equivalent network parameters are employed to describe the antenna behavior within the context of an electromagnetic circuit. Accordingly, the antenna may be replaced by equivalent circuit models such as those given in Figure 1-6, where the antenna is characterized by a) its input impedance Z and a source voltage V_0 or b) its input admittance Y and a source current I_0 . In the transmitting mode, V_0 or I_0 is zero, and V_{in} or I_{in} is the input voltage or current that drives the antenna. Part of the source power is dissipated in the loss resistance R_{loss} or loss conductance G_{loss} . The power associated with the radiation resistance R_{rad} or radiation conductance G_{rad} is responsible for the radiated wave that escapes from the antenna to the surrounding space. The quantities X and B refer to the equivalent reactance and susceptance of the antenna.

When the antenna is receiving, the terminals of the equivalent circuit are connected to a load (amplifier, speaker, etc.), and V_0 or I_0 are non-zero as dictated by the external wave that impinges on the antenna. Maximum transmission to the load occurs when the load impedance is the conjugate of the input antenna impedance. Some of the received energy usually remains within the radiation resistance or conductance and is re-radiated to contribute to the radar cross section (RCS) of the antenna. Mismatches between the load and the antenna input impedance can lead to higher or lower RCS. However, care must be exercised when

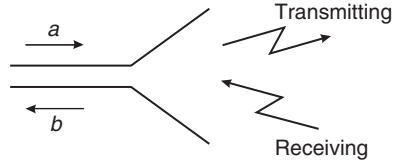


FIGURE 1-5 Transition between waveguide (transmission line) and antenna

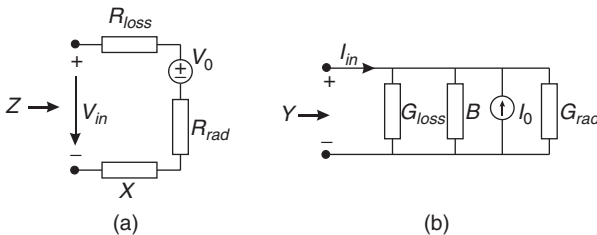


FIGURE 1-6 (a) Impedance (Thevenin equivalent) and (b) admittance (Norton equivalent) representations of an antenna

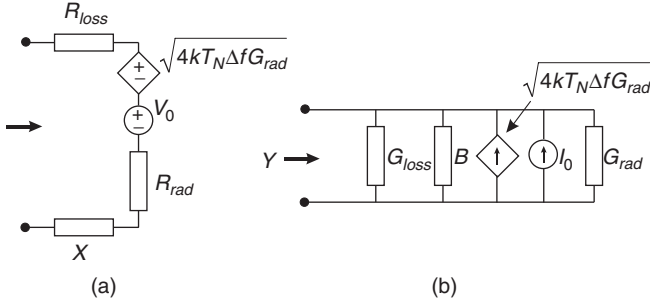


FIGURE 1-7 (a) Impedance (Thevenin equivalent) and (b) admittance (Norton equivalent) representations of an antenna including noise sources

the re-radiated energy is determined using equivalent circuit representations such as those in Figure 1-6. Reasonable results can only be expected if the employed equivalent circuit is an appropriate representation of the antenna's internal structure. Also, the equivalent circuits of one antenna can be different for radiation and reception.

Each antenna receives noise radiation related to the brightness temperature of its environment. Usually, the received noise power restricts the performance of an antenna and subsequent processing devices. However, in remote sensing or astronomy applications, the noise power can even be the intended signal. To formally account for the received noise power, the equivalent circuits in Figure 1-6 must be used with V_0 and I_0 representing sources dependent on the equivalent noise temperature T_N of the antenna. The dependent source voltages and source currents are given in Figure 1-7, where k is the Boltzmann constant and $BW = \Delta f$ is the receiver or antenna bandwidth. T_N is dependent on the radiation (or rather receiving) pattern and orientation of the antenna as well as the distribution of the brightness temperature in the environment of the antenna.[†]

Usually, antennas work at certain frequencies. The bandwidth

$$BW = f_U - f_L \quad (1-17)$$

is the operational frequency range of the antenna around some center frequency

$$f_0 = 1/2(f_U - f_L) \quad (1-18)$$

(e.g., the resonance frequency of the antenna). In this equation, the upper operational frequency limit is f_U , and the lower operational frequency is f_L . Often, the bandwidth is quoted with respect to f_0 percent, and in that case, it is given by

$$BW = (f_U - f_L)/f_0 \text{ } 100\% \quad (1-19)$$

From a circuit point of view, the matching of an antenna to the generator is very important. Therefore, bandwidth definitions with respect to the reflection coefficients or input impedances are common, for example,

$$|\Gamma| = |(Z_{in} - Z_0)/(Z_{in} + Z_0)| < 0.5 \quad (1-20)$$

[†] Further noise considerations with respect to antennas can be found in the literature.^{1,2,11,16}

Typically, the voltage standing wave ratio (VSWR)

$$\text{VSWR} = (|\Gamma| + 1)/(|\Gamma| - 1) < 2 \quad (1-21)$$

is used to define the operational bandwidth. Bandwidth definitions with respect to gain or other antenna pattern characteristics (for example, $G_0/G_{0,\max} < 0.5$) can be used to characterize the antenna radiating and receiving properties.

1.6 POLARIZATION

Polarization is a property of a single-frequency electromagnetic wave; it describes the shape and orientation of the locus of the extremity of the field vectors as a function of time.^{7,8} In antenna engineering, the polarization properties of plane waves or waves that can be considered to be planar over the local region of observation are of interest. For plane waves, it is sufficient to specify the polarization properties of the electric field vector since the magnetic field vector is simply related to the electric field vector. The plane containing the electric and magnetic fields is called the *plane of polarization* and is orthogonal to the direction of propagation. Generally, the tip of the electric field vector moves along an elliptical path in the plane of polarization. The polarization of the wave is specified by the shape and orientation of the ellipse and the direction in which the electric field vector traverses the ellipse.

The shape of the ellipse is specified by its *axial ratio*—the ratio of the major axis to the minor axis. Its orientation is specified by the *tilt angle*—the angle between the major axis and a reference direction when viewed looking in the direction of propagation. The direction in which the electric field vector traverses the ellipse is the *sense of polarization*—right-handed or left-handed when viewed looking in the direction of propagation.

The polarization of an antenna in a specific direction is defined to be the polarization of the far-field radiated in that direction. Usually, the polarization of an antenna remains relatively constant throughout the main lobe, but varies considerably in the minor lobes. It is convenient to define a spherical coordinate system associated with an antenna as illustrated in Figure 1-8. The polarization ellipse for the direction (θ, ϕ) is shown inscribed on

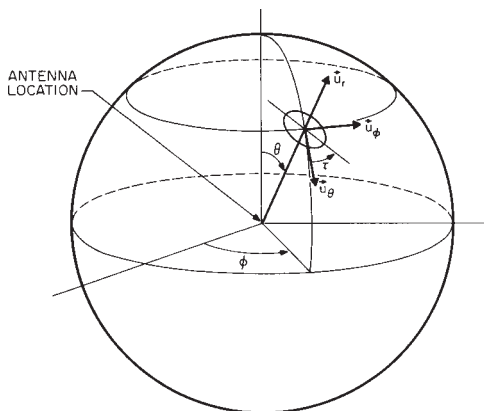


FIGURE 1-8 Polarization ellipse in relation to the antenna coordinate system (after⁷ © IEEE 1979)

the spherical shell surrounding the antenna. It is common practice to choose \mathbf{u}_θ (the unit vector in the θ direction) as the reference direction. The tilt angle is then measured from \mathbf{u}_θ toward \mathbf{u}_ϕ . The sense of polarization is clockwise if the electric field vector traverses the ellipse from \mathbf{u}_θ toward \mathbf{u}_ϕ , as viewed in the direction of propagation and counterclockwise if the reverse is true.

In many practical situations, such as antenna measurements, it is convenient to establish a local coordinate system. Usually, the u_3 axis is the direction of propagation; the u_1 axis is horizontal; and the u_2 axis is orthogonal to the other two so the unit vectors are related by $\mathbf{u}_1 \times \mathbf{u}_2 = \mathbf{u}_3$. The tilt angle is measured from \mathbf{u}_1 . When an antenna receives a wave from a particular direction, the response will be greatest if the polarization of the incident wave has the same axial ratio, the same sense of polarization, and the same spatial orientation as the polarization of the antenna in that direction. This situation is depicted in Figure 1-9, where E_t represents a transmitted wave (antenna polarization) and E_m represents a matched incident wave. Note that the sense of polarization for E_t and E_m are the same when viewed in their local coordinate system. Also, note that the tilt angles are different because the directions of propagation are opposite. As depicted in Figure 1-9, τ_t is the tilt angle of the transmitted wave and τ_m is the tilt angle of the polarization-matched received wave; they are related by

$$\tau_m = 180^\circ - \tau_t \quad (1-22)$$

The polarization of the matched incident wave, as just described, is called the *receiving polarization* of the antenna.

When the polarization of the incident wave is different from the receiving polarization of the antenna, then a loss due to polarization mismatch occurs. Let

$$\eta_p = \text{polarization efficiency (unitless)}$$

where η_p is the ratio of the power received by the antenna to the power received when polarization of the incident wave is matched to the receiving antenna polarization.

The Poincaré sphere, shown in Figure 1-10, is a convenient representation of polarization states. Each possible polarization state is represented by a unique point on the unit sphere. Latitude represents axial ratio, with the poles being circular polarizations; the upper hemisphere is for left-handed sense, and the lower hemisphere is for right-handed sense. Longitude represents tilt angles from 0 to 180° . An interesting feature of the Poincaré sphere is that diametrically opposite points represent orthogonal polarizations.

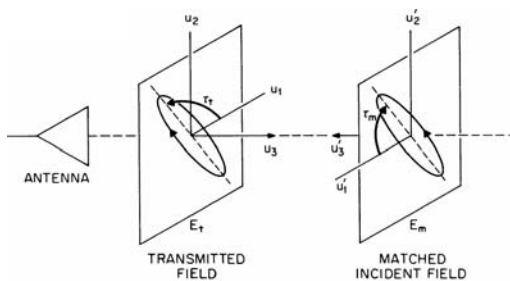


FIGURE 1-9 Relation between polarization properties of an antenna when transmitting and receiving (after⁷ © IEEE 1979)

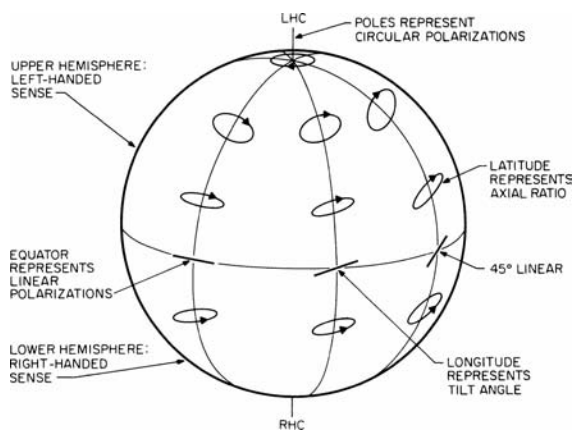


FIGURE 1-10 Polarization states on the Poincaré sphere (after⁷ © IEEE 1979)

The Poincaré sphere is also convenient for representing polarization efficiency. In Figure 1-11, W represents the polarization of an incident wave, and A_r is the receiving antenna polarization. If the angular distance between the points is 2ξ , then the polarization efficiency is

$$\eta_p = \cos^2 \xi \quad (1-23)$$

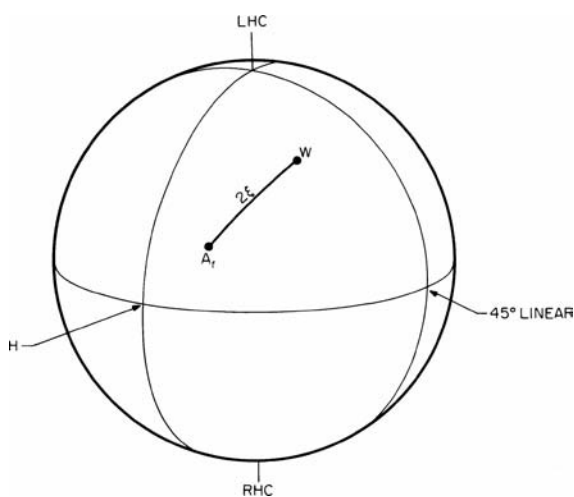


FIGURE 1-11 Receiving polarization of an antenna A_r for an incident wave polarization W

1.7 DIRECTIVITY PATTERNS FROM CONTINUOUS LINE SOURCES

According to Huygens' and the equivalence principle (see Section 1.2), the radiation of arbitrary antennas can be characterized by considering equivalent source current distributions. The simplest source distributions are the electric and magnetic point sources considered in Section 1.3. More degrees of freedom toward the realization of particular directivity properties are provided by continuous line sources that can, for instance, be used to characterize the radiation behavior of linear wire antennas.

For line sources, the current distribution (electric and/or magnetic) is considered to be a function of only a single coordinate. The directivity pattern $E(u)$ resulting from a given distribution is simply related to the distribution by a finite Fourier transform,^{5,9,10} given by

$$E(u) = \frac{\ell}{2} \int_{-1}^{+1} f(x) e^{jux} dx, \quad (1-24)$$

where

$f(x)$ = relative shape of source distribution over aperture as a function of x

$u = (\pi\ell\lambda) \sin \phi$

ℓ = overall length of aperture

ϕ = angle measured from normal to aperture

x = normalized distance from normal to aperture $-1 \leq x \leq 1$

The simplest type of line source distribution is the uniform distribution where $f(x) = 1$ along the aperture and is zero elsewhere. The directivity pattern for such an antenna is (see Chapter 3)



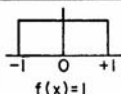
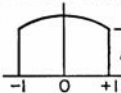
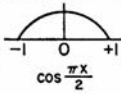
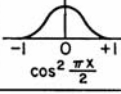
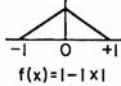
$$E(u) = \ell \frac{\sin u}{u} = \ell \frac{\left[\left(\frac{\pi\ell}{\lambda} \right) \sin \phi \right]}{\left(\frac{\pi\ell}{\lambda} \right) \sin \phi} \quad (1-25)$$

This type of directivity pattern $20 \log_{10} |E(u)|$ is of interest because all field contributions add in phase, giving the highest gain without cancellation effects.⁵ However, sidelobe levels are high, and the intensity of the first sidelobe is only 13.2 dB less than the maximum.

The intensity of the sidelobe levels can be reduced considerably by tapering the aperture distribution so the amplitude drops off smoothly from the center of the aperture to its edges. There are an unlimited number of possible distributions. However, a few simple types of distributions are typical and illustrate how the beamwidth, sidelobe level, and relative gain vary as a function of the distribution. Table 1-1 gives the important characteristics of several distributions having a simple mathematical form.

Of considerable interest is the manner in which the sidelobes fall off as the angle from the main beam increases or as u increases. For the uniform distribution that has a discontinuity in both the function and its derivatives at the edge of the aperture, the sidelobes decrease as u^{-1} . For the gable distribution or cosine distribution, both of which are continuous at the edge of the aperture but that have discontinuous first derivatives, the far-out sidelobes fall off as u^{-2} . For the cosine-squared distribution that has a discontinuous second derivative, the far-out sidelobes fall off as u^{-3} . Many distributions (obtained in practice) can be approximated by one of simpler forms or by a combination of simple forms.

TABLE 1-1 Line-Source Distributions

TYPE OF DISTRIBUTION $-l \leq x \leq l$	DIRECTIVITY PATTERN $E(u)$	HALF POWER BEAMWIDTH IN DEGREES 	ANGULAR DISTANCE TO FIRST ZERO 	INTENSITY OF 1st SIDELobe db BELOW MAX.	GAIN FACTOR
 $f(x) = 1$	$\ell \frac{\sin u}{u}$	$50.8 \frac{\lambda}{\ell}$	$57.3 \frac{\lambda}{\ell}$	13.2	1.0
 $f(x) = 1 - (1 - \Delta)x^2$	$\ell (1 + \Delta) \frac{\sin u}{u}$ $\Delta = (1 - \Delta) \frac{d^2}{du^2}$	$\Delta = 1.0$	$50.8 \frac{\lambda}{\ell}$	13.2	1.0
		.8	$52.7 \frac{\lambda}{\ell}$	15.8	.994
		.5	$55.6 \frac{\lambda}{\ell}$	17.1	.970
		0	$65.9 \frac{\lambda}{\ell}$	20.6	.833
 $\cos \frac{\pi x}{2}$	$\frac{\pi \ell}{2} \frac{\cos u}{(\frac{\pi}{2})^2 - u^2}$	$68.8 \frac{\lambda}{\ell}$	$85.9 \frac{\lambda}{\ell}$	23	.810
 $\cos^2 \frac{\pi x}{2}$	$\frac{\ell}{2} \frac{\sin u}{u} \frac{\pi^2}{\pi^2 - u^2}$	$83.2 \frac{\lambda}{\ell}$	$114.6 \frac{\lambda}{\ell}$	32	.667
 $f(x) = 1 - x $	$\frac{\ell}{2} \left(\frac{\sin \frac{u}{2}}{\frac{u}{2}} \right)^2$	$73.4 \frac{\lambda}{\ell}$	$114.6 \frac{\lambda}{\ell}$	26.4	.75

For instance, suppose you wanted to find the directivity pattern of a cosine-squared distribution on a pedestal, i.e., a combination of a uniform distribution and a cosine-squared distribution given by

$$f(x) = C + \cos^2 \frac{\pi x}{2} \quad (1-26)$$

The resulting directivity pattern is then obtained directly by superposition to obtain the direction pattern:

$$E(u) = C \ell \frac{\sin u}{u} + \frac{\ell}{2} \frac{\sin u}{u} \frac{\pi^2}{\pi^2 - u^2} \quad (1-27)$$

Note that the sidelobes and other characteristics of the pattern must be obtained from the new directivity pattern and cannot be interpolated from Table 1-1. By choosing the proper relative intensities of a uniform distribution and a cosine-squared distribution, it is possible to obtain a theoretical sidelobe level that is very low. For instance, if $C = 0.071$, then the intensity of the largest sidelobe will be 43 dB less than the maximum of the main beam with a half-power beamwidth given by $76.5\lambda/\ell$, a value that is somewhat lower than that of the cosine-squared distribution by itself.

In practice, it is not easy to synthesize prespecified continuous line-source distributions. Consider, for instance, a linear wire antenna; the electric current distribution along the wire is determined to fulfill Maxwell's equations under the constraint of the given boundary

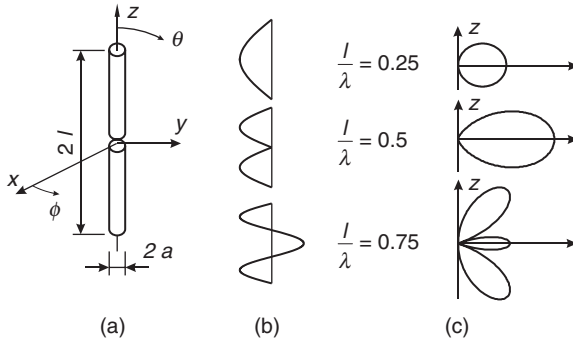


FIGURE 1-12 (a) Linear wire antenna geometry, (b) Current distributions according to Eq. 1-28, and (c) corresponding field radiation patterns in arbitrary $\phi = \text{const.}$ planes (linear scale) for the first three resonances

conditions and cannot, therefore, be shaped arbitrarily. Good approximation of the current can be done, however, to study radiation from specific antennas.

In the following considerations, focus will be on a linear center-fed dipole wire antenna of length $2l$ aligned along the z axis, as illustrated in Figure 1-12a. The characteristics of the corresponding monopole antenna of length l over a perfectly conducting plane can be obtained by applying image theory.^{1,3}

As a first approximation, the thin wire antenna can be replaced by the z directed current distribution

$$I(z) = I_0 \sin(k(l - |z|)) \quad (1-28)$$

defined along the wire length. The corresponding electric field intensity in the far-field is³

$$\mathbf{E}(r) = \frac{jI_0}{2\pi r} \frac{\mu_0}{\epsilon_0} e^{-jk_0 r} \left[\frac{\cos(k_0 l \cos \theta) - \cos(k_0 l)}{\sin \theta} \right] \hat{\theta} \quad (1-29)$$

and can be used to derive an approximate far-field pattern. Obviously, the pattern is omnidirectional, and its θ dependence varies with antenna length. The θ dependence (in linear scale) of the field radiation pattern is illustrated in Figure 1-12c for three different wire lengths, with the corresponding current distributions according to Eq. 1-28 as depicted in Figure 1-12b. The radiation resistance can be calculated by integrating the total radiated power and relating it to the appropriate input current at the antenna port.^{1,3} Evaluating this expression gives the resistance as depicted in Figure 1-13.

Wire antennas are usually operated close to their resonance lengths $2l/\lambda = 0.5, 1, 1.5, 2, \dots$, where $2l/\lambda = 0.5, 1.5, \dots$ corresponds to the current resonance with low radiation resistance. Lengths $2l/\lambda = 1, 2, \dots$ result in voltage resonance with high radiation resistance. The lowest-order resonance $2l/\lambda = 0.5$ is associated with a radiation resistance of approximately 73Ω . Difficulties due to nulls in the current distributions for voltage resonances can be overcome by a modified higher-order current distribution.^{‡1} Figure 1-14

[‡] Extensive studies of linear wire antennas, even with realistic geometric dimensions, are presented in R. W. P. King's *The Theory of Linear Antennas*.¹² Approximate integral equation solutions are the basis for characterizing the various antenna parameters and results for a large variety of parameters are given.

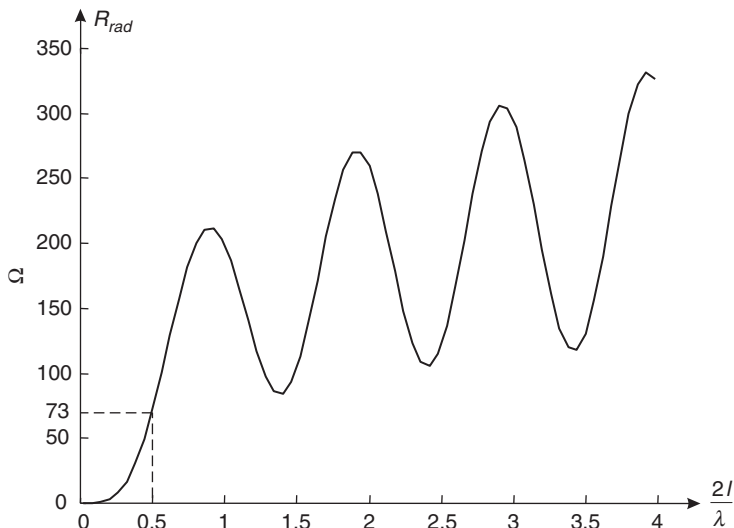


FIGURE 1-13 Wire antenna radiation resistance obtained from the current distribution of Eq. 1-28 (normalized to I_0)

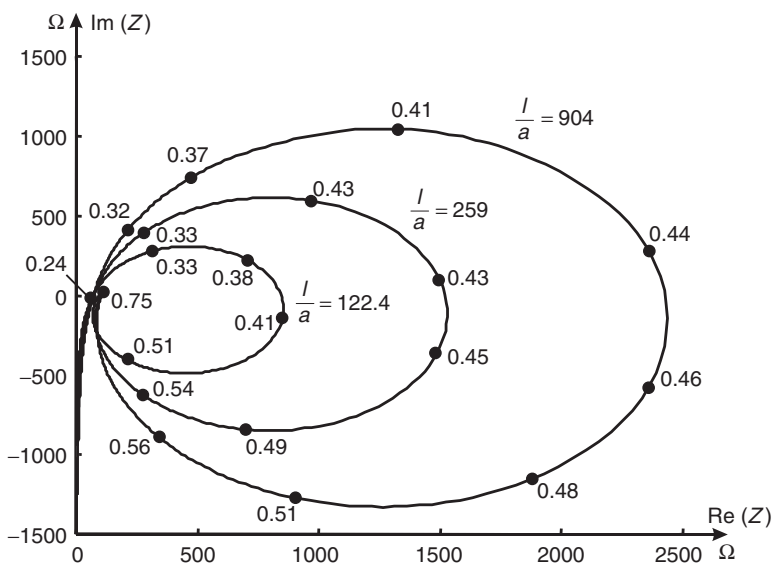


FIGURE 1-14 King-Middleton¹² second-order impedances for different values of l/a (see Figure 1-12). The numbers given next to the dots are the l/λ values.

illustrates the complex input impedance of the wire antenna for different l/a ratios and l/λ values.[§] You can see that thicker antennas have smaller impedance oscillations resulting in larger bandwidths. Also, the impedance curves show a slight shift of the reactance to the capacitive range.

A large collection of numerical computer codes, often based on the method of moments^{1,11} (see Chapter 59 on numerical techniques), is now available for analysis, design, and optimization of many wire antenna types.

1.8 DIRECTIVITY PATTERNS FROM AREA SOURCE DISTRIBUTIONS

Area source distributions serve as a radiation model for many antenna types, especially for those with radiation in a particular direction. Insight into various antenna characteristics can be gained from the consideration of simple aperture shapes such as rectangular, circular, or elliptical apertures.

Rectangular Apertures

The directivity pattern of an area distribution is found in a similar manner to that used for line-source distributions, except the aperture field is integrated over two dimensions instead of one dimension. If the aperture distribution is given by $f(x, y)$, where x and y are the two coordinates, then the directivity pattern is given by

$$E(\theta, \phi) = \iint f(x, y) e^{jk_0 \sin \theta (x \cos \phi + y \sin \phi)} dx dy \quad (1-30)$$

The difficulty of evaluating this expression depends on the form of the distribution function. For many antenna types, such as the rectangular horn, for example, the distribution function is separable, that is

$$f(x, y) = f(x)f(y)$$

The directivity patterns in the principal planes are readily determined for the separable case because the pattern in the xz plane is identical to the pattern produced by a line-source distribution $f(x)$, whereas the pattern in the yz plane is identical to the pattern produced by a line-source distribution $f(y)$. If the distribution function is not separable, the integral must be evaluated either analytically, graphically, or numerically.

Circular Apertures

An antenna that is used frequently in microwave applications is a *paraboloid* having circular symmetry. The radiation pattern can be computed by projecting the field distribution

[§] The curves in Figure 1-14 represent the King-Middleton second-order impedances and are drawn from the numerical data in *The Theory of Linear Antennas*.¹²

on the paraboloid to a plane at the opening of the paraboloid and computing the directivity pattern due to the plane aperture:

$$E(u, \phi') = a^2 \int_0^{2\pi} \int_0^1 f(r, \phi') e^{j u r \cos(\phi - \phi')} r dr d\phi' \quad (1-31)$$

where

a = radius at outside of aperture

p = radius at any point inside aperture

$r = p/a$

$u = (2\pi a/\lambda) \sin \theta = (\pi D/\lambda) \sin \theta$

$D = 2a$ = aperture diameter

$f(r, \phi')$ = the normalized aperture distribution function

The coordinates are shown in Figure 1-15.

The simplest forms of aperture distributions to evaluate are those in which the distribution is independent of the angular coordinate ϕ' but depends only on the radial coordinate r . The integral for the directivity pattern then becomes

$$E(u) = 2\pi a^2 \int_0^1 f(r) J_0(ur) r dr \quad (1-32)$$

When the distribution is constant, the integral is evaluated to give

$$E(u) = 2\pi a^2 \frac{J_1(u)}{u} \quad (1-33)$$

Antenna engineers frequently need to evaluate the directivity pattern for an illumination that tapers down toward the edge of the aperture. One function, which is convenient for representing the aperture distribution, is

$$f(r) = (1 - r^2)^p \quad (1-34)$$

This function behaves in a similar fashion to the n th-power distributions, as discussed for the line-source case (Section 1.7). When the exponent increases, the distribution becomes more tapered and more concentrated to the center of the aperture. When the exponent decreases and approaches zero, the distribution approaches uniform illumination.

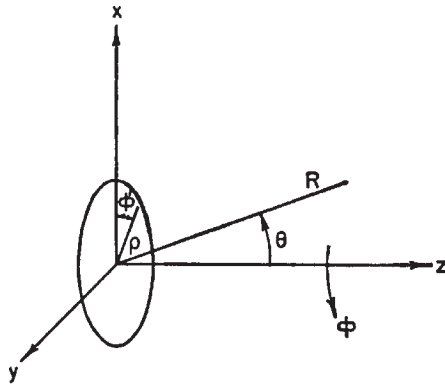


FIGURE 1-15 Coordinates for a circular aperture

Evaluating the directivity pattern, gives

$$E(u) = 2\pi a^2 \int_0^1 (1-r^2)^p J_0(ur) dr$$

$$= \pi a^2 \frac{2^p p! J_{p+1}(u)}{u^{p+1}} = \frac{a^2}{p+1} \Delta_{p+1}(u), \quad (1-35)$$

where the Bessel functions $J_{p+1}(u)$ and the lambda function $\Delta_{p+1}(u)$ are available in tabular form.¹³

The principal characteristics of the directivity patterns are given in Table 1-2 for the cases $p = 0, 1, 2$. Comparison of the patterns for the uniformly illuminated circular aperture (i.e., when $p = 0$) with results for the uniformly illuminated line source (Section 1.7) shows that the circular aperture has a lower sidelobe level and a broader beamwidth. This is expected because projections of the circular-aperture illumination onto a line produce an equivalent line source that is no longer uniform but has some degree of tapering.



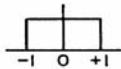

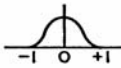
Elliptical Apertures

In some applications, an elliptically shaped reflector is used to permit control of the relative beamwidth in the two principal planes and to control the sidelobes by shaping the reflector outline. Computation of the directivity patterns for the aperture shape can be carried out by knowing the Fourier components of the illumination function over the aperture.

Realization of Continuous Aperture Distributions

In practice, continuous aperture distributions of large extent are realized by *horn, reflector, or lens antennas*.

TABLE 1-2 Circular-Aperture Distributions

TYPE OF DISTRIBUTION $0 \leq r \leq 1$	DIRECTIVITY PATTERN $E(u)$	HALF POWER BEAMWIDTH IN DEGREES 	ANGULAR DISTANCE TO FIRST ZERO 	INTENSITY OF 1st SIDELOBE db BELOW MAX.	GAIN FACTOR
 $f(r) = (1-r^2)^0 = 1$	$\pi a^2 \frac{J_1(u)}{u}$	$58.9 \frac{\lambda}{D}$	$69.8 \frac{\lambda}{D}$	17.6	1.00
 $f(r) = (1-r^2)$	$2\pi a^2 \frac{J_2(u)}{u^2}$	$72.7 \frac{\lambda}{D}$	$93.6 \frac{\lambda}{D}$	24.6	0.75
 $f(r) = (1-r^2)^2$	$8\pi a^2 \frac{J_3(u)}{u^3}$	$84.3 \frac{\lambda}{D}$	$116.2 \frac{\lambda}{D}$	30.6	0.56

Horn antennas (see Figure 1-16) are among the oldest and most popular microwave antennas. They can deliver 10 to 30 dB directivity, and their robust metallic constructions and waveguide feeds permit high-power handling. Horns are, therefore, often used as feeds for reflector antennas realizing even larger apertures. Horns are extensions of the widely used microwave guiding devices, such as rectangular and circular hollow waveguides in a natural manner (see Figure 1-16). To achieve matching from the waveguide to free-space, the hollow waveguide is tapered to a larger opening called the *aperture*. The effective aperture of the antenna (Section 1.4) is almost identical to the geometric aperture of the horn.

The equivalent aperture source distributions are typically used for extracting the radiation properties of horn antennas. These sources are an approximation found from the guided wave field arriving at the horn's aperture. Smaller beamwidths and larger directivities can be expected for horns with larger cross-section apertures, a consequence of the Fourier transformation relation between aperture distributions and far-field patterns. However, the maximum achievable directivity is limited due to phase errors of the aperture source distributions caused by the curved phase fronts in the tapered part of the horn antenna. Phase error corrections can be realized by using lenses in the horn aperture or by connecting the horn to a parabolic reflector (see Figures 1-17, 1-18, 1-19).

Another important concept in designing horn antennas is *corrugation*. Corrugations are grooves on the waveguide walls and are equivalent to introducing anisotropic boundary conditions on these walls. These anisotropic conditions offer additional degrees of freedom for controlling the radiation pattern and have been shown to lead to lower cross polarizations, higher aperture efficiencies, and more symmetrical far-field patterns.

The E plane sectoral horn is tapered only in the plane containing the E field (of the lowest-order rectangular waveguide mode), and the H plane sectoral horn is tapered in the corresponding plane containing the H field. Correspondingly, the pyramidal horn is tapered in both planes, and horns that are extensions of a circular waveguide are referred to as conical horn antennas. Further details on the analysis and practical realization of horn antennas can be found in Chapter 14 and the literature.^{1,2,14}

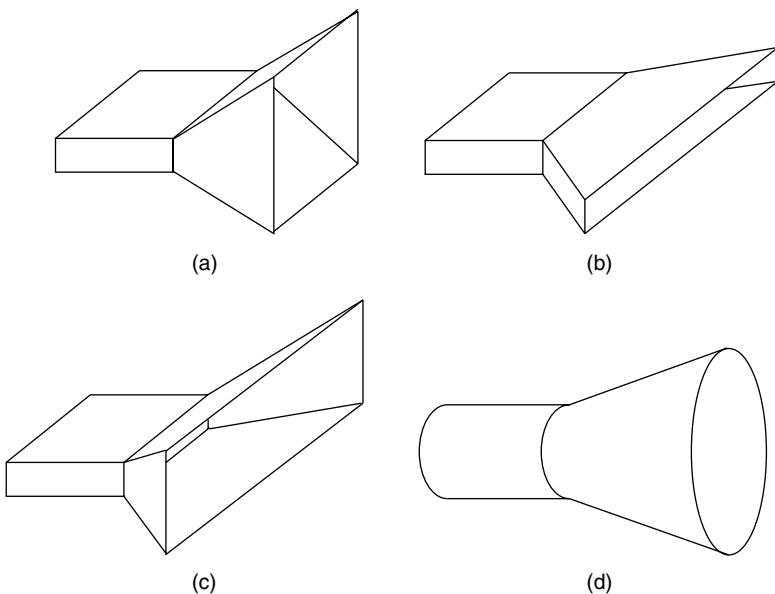


FIGURE 1-16 Common electromagnetic horn antennas: (a) E plane horn, (b) H plane horn, (c) pyramidal horn, (d) conical horn

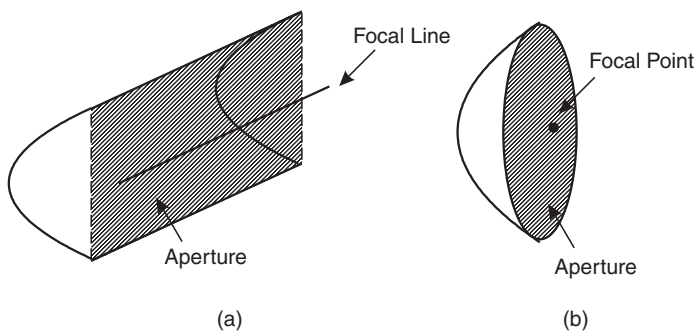


FIGURE 1-17 (a) Cylindrical parabolic reflector with focal line and (b) paraboloidal reflector with focal point

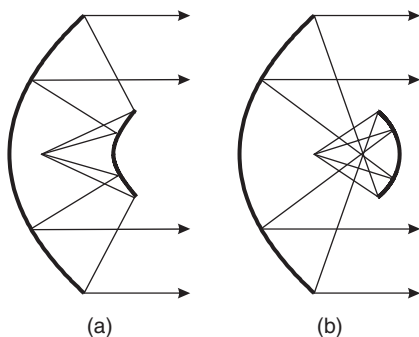


FIGURE 1-18 (a) Cassegrain and (b) Gregorian double-reflector systems

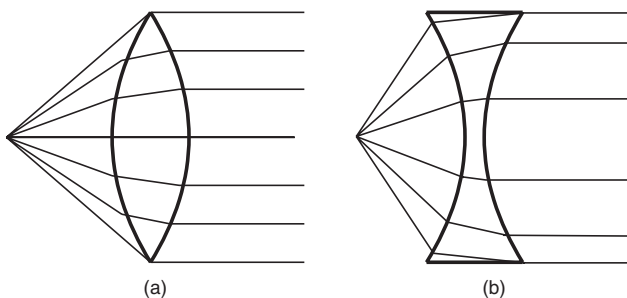


FIGURE 1-19 (a) Delay lens with index of refraction > 1 and (b) fast lens with index of refraction < 1

From the preceding paragraphs and especially with knowledge of the Fourier transform relation between far-field patterns and aperture distributions, you can observe that high-gain antennas with small beamwidths typically require large (equivalent) aperture source distributions. Even though large apertures can be realized, however, unavoidable phase errors, as in the case of horn antennas, restrict achievable gain and correspondingly the effective aperture. A way to overcome this issue is to use a secondary device for shaping the phase fronts of the radiated or received waves. Metallic reflectors (curved metallic surfaces) serve this purpose and have been used since the early days of antennas to realize high gains. There are many different reflector types² ranging from linear reflector elements over flat metallic sheets to more complicated shapes such as corner reflectors, circular reflectors, hyperbolic reflectors, elliptic reflectors, and the more commonly used parabolic reflectors. Next, we briefly discuss parabolic reflectors. For other reflector types, refer to Chapter 15 on reflector antennas and the literature.^{1,2,15}

Figure 1-17 depicts the cylindrical and spherical parabolic reflectors. Both utilize the unique feature of the parabola to adjust the ray path lengths so that spherical waves emanating from the focal point generate fields of the same phase on the reflector aperture after reflection from the reflector surface. As a result, the rotationally symmetric paraboloid gives rise to very narrow patterns. Highly directive antennas can be realized by placing a primary line source in the focal line of a cylindrical paraboloidal reflector or a point source (a horn) at the focus of a paraboloid. Of course, these primary sources should be directed toward the reflector. Because the feed causes undesirable blockage and distortions to the main beam (a situation that worsens due to collateral equipment around the feed), subreflectors are introduced to place the feed at different locations. Shown in Figure 1-18 are classical *Cassegrain* and *Gregorian* double-reflector arrangements illustrating this concept. Cassegrain antennas use a hyperbolic subreflector whereas Gregorian reflectors employ an elliptical subreflector. Both subreflector types are usually designed so the feed or new focus can be located close to the primary reflector permitting feeding of the source at the focus through a hole in the primary reflector. Electromagnetic analysis of reflector antennas usually follows the principles of geometrical and physical optics, or geometrical theory of diffraction^{1,2,11,16} to account for edge diffraction.

Two different analysis techniques are often employed for reflector analysis: the aperture distribution and current distribution methods. For the latter, the electric currents are found directly on the reflector dishes, and the radiated fields are calculated by integrating these currents (see Section 1.3). In the aperture distribution method, equivalent currents are derived and placed on a planar aperture in front of the reflector, as illustrated in Figure 1-17. The radiated fields are then found by integrating these equivalent currents. As noted earlier, aperture blocking by the primary feed or the subreflector reduces antenna efficiency. Also, the blockage gives rise to higher sidelobe levels and cross-polarization effects. These blockages can be circumvented by reshaping the reflector so the feed is offset from the center of the aperture,¹⁷ often below or outside of the aperture view.

Similar to reflector antennas, *lens antennas* provide a means of shaping phasefronts of electromagnetic waves and influencing wave propagation in certain directions. Lens antennas can be divided into *delay lenses* and *fast lenses* (see Figure 1-19).² In a delay lens medium, the electromagnetic path length is increased (refractive index $n > 1$), whereas in a fast lens medium, the electromagnetic path length is decreased inside the lens (refractive index $n < 1$). Lenses may also be divided among *dielectric* and *metal-plate* lenses. Dielectric lenses consist of natural dielectrics or artificial dielectrics. Metal-plate lenses consist of parallel metallic plates realizing parallel-plate waveguides between the individual plates. *H* plane metal-plate lenses have metallic plates parallel to the *H* field of the electromagnetic wave and lead to $n > 1$. In contrast, the *E* field metal-plate lenses have the plates parallel to the *E* field resulting in $n < 1$. As noted earlier, lens antennas are often used to form collimating beams and thus increase the gain of microwave antennas. In this context, their purpose is to transform spherical or cylindrical wavefronts into planar wavefronts, and this is the reverse of focusing plane waves into a focal line or point (see Figure 1-19).

In contrast to reflector antennas, lens antennas do not have difficulties with aperture blocking and allow for beam scanning over relatively large angles. They also provide for additional design freedom since wave propagation is influenced by refraction at both the surfaces and the refractive index inside the medium, which can be adjusted as required (for instance by using layered designs). However, the corresponding design process is more complex when compared to metallic reflector antennas. Difficulties with lenses are caused by reflections at the two surfaces and losses in the lens material. Nevertheless, surface reflections can be reduced by matching layers or surface roughness. Analysis and design procedures for lens antennas are typically based on geometrical optics (Fermat's principle), physical optics, and related approaches such as the geometrical theory of diffraction. Details on the analysis of lens antennas and on the various classes of such antennas can be found in the literature^{1,2,11,16} and in Chapter 18 on lens antennas.

Recently, ideal lenses with arbitrarily good focusing properties have been discussed. Such lenses should be realizable by the use of so-called metamaterials with a refractive index of -1 .¹⁸

1.9 FUNDAMENTALS OF ANTENNA ARRAYS

The radiation and receiving characteristics of antennas can be shaped by synthesizing certain equivalent source distributions. Because it is difficult to control continuous aperture currents or fields, discrete configurations are often used, leading to the concept of *antenna arrays*. Putting the elements of an antenna array in a certain pattern and adjusting the amplitude and phase of the individual antenna elements appropriately allows for the synthesis of arbitrary aperture sources. These behave quite similarly to continuous aperture distributions (provided certain rules with respect to element spacing are followed). Design difficulties often arise due to coupling among array elements. Consequently, the amplitude and phase of the individual array elements cannot be adjusted independently from one another. The driving-point impedance of an individual element might differ considerably from its self-impedance because of the mutual coupling with other array elements. In a multi-element array, a way to relate the terminal voltages and element currents is

$$\begin{aligned} V_1 &= I_1 Z_{11} + I_2 Z_{12} + \cdots + I_n Z_{1n} \\ V_2 &= I_1 Z_{12} + I_2 Z_{22} + \cdots + I_n Z_{2n} \\ &\vdots \\ &\vdots \\ V_n &= I_1 Z_{1n} + I_2 Z_{2n} + \cdots + I_n Z_{nn} \end{aligned} \quad (1-36)$$

where

- V_n = impressed voltage at the n th element
- I_n = current flowing in the n th element
- Z_{nn} = self-impedance of the n th element
- $Z_{mn} = Z_{nm}$ = mutual impedance between m th and n th elements

The driving-point impedance for element 1 is found from the ratio of the impressed voltage to the current and is obtained from the previous equation as follows:

$$Z_{1\text{input}} = \frac{V_1}{I_1} = Z_{11} + \frac{I_2}{I_1} Z_{12} + \cdots + \frac{I_n}{I_1} Z_{1n} \quad (1-37)$$

The reader can see that the input or driving-point impedance of a particular element is not only a function of its own self-impedance but also a function of the relative currents

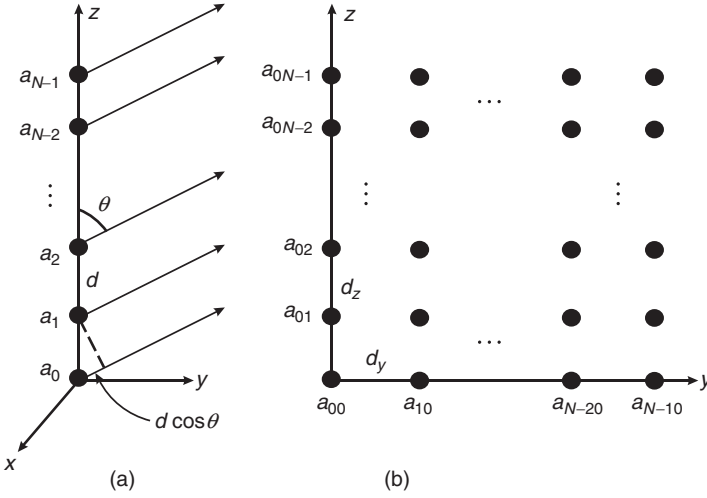


FIGURE 1-20 (a) Equally spaced linear and (b) two-dimensional array configurations

flowing on the other elements and of mutual impedance between elements. In an array in which the current distribution is critical, it is necessary to determine the input impedance from the previously described relationship and to design the feed system to match the input impedance rather than the self-impedance.**

To account for element coupling in the design, rigorous numerical methods^{1,4,11} must be employed. However, for simplicity, in the following array, coupling will be neglected because the main focus is the introduction of some basic concepts in array design.

A very basic and important array configuration is that of a equally spaced linear array of N identical elements on a straight line, as illustrated in Figure 1-20(a). The element spacing is d , and a linear phase progression is assumed for the element excitation currents. The total electric field intensity \mathbf{E}_{tot} in the far-field is given by

$$\begin{aligned} \mathbf{E}_{\text{tot}}(\theta, \phi) &= \mathbf{E}_{el}(\theta, \phi) \sum_{n=0}^{N-1} a_n e^{jn(k_0 d \cos \theta - \alpha)} \\ &= \mathbf{E}_{el}(\theta, \phi) \sum_{n=0}^{N-1} a_n e^{jn\psi} = \mathbf{E}_{el}(\theta, \phi) f(\psi) \end{aligned} \quad (1-38)$$

where $\psi = k_0 d \cos \theta - \alpha$ and \mathbf{E}_{el} refers to the array element pattern. Also, a_n is the amplitude of the individual array elements; α is the phase progression from one element to the next; and k_0 is the wavenumber of free-space. If all excitation currents are equal in amplitude ($a_0 = a_1 = a_2 = \dots = a_{N-1}$), the array factor $f(\psi)$ becomes

$$f(\psi) = a_0 \sum_{n=0}^{N-1} e^{jn\psi} = a_0 \frac{1 - e^{jN\psi}}{1 - e^{j\psi}} \quad (1-39)$$

** Some examples of this are given in "The Effect of a Periodic Variation in the Field Intensity Across a Radiating Aperture."¹⁹ See Chapter 3 for more information.

This can be simplified to obtain the normalized form

$$f_0(\psi) = \frac{f(\psi)}{Na_0} = \frac{\sin \frac{N\psi}{2}}{N \sin \frac{\psi}{2}} \quad (1-40)$$

where $f_0(\psi)$ is maximum when $\psi = 0$. For broadside radiation, $\alpha = 0$ must hold, implying that $\theta = \pi/2$. To scan the array beam toward other directions, α must be selected so that $k_0 d \cos \theta = \alpha$, resulting in

$$\theta = \theta_{\max} = \cos^{-1} \left(\frac{\alpha}{k_0 d} \right) \quad (1-41)$$

as the angle of the maximum array radiation. The direction of maximum radiation can be adjusted by controlling α , and this is a concept characteristic to scanning arrays. In practical phased arrays, the phase of the individual array elements is controlled electronically to allow for much more flexible, faster, and reliable array scanning as compared to traditional mechanical steering approaches.

Modern mobile communication systems are increasingly employing phased arrays at base stations (in conjunction with sophisticated signal processing algorithms) to expand the base station customer capacity and reduce interference among adjacent stations. In the wireless industry, such antennas are typically referred to as *smart* or *adaptive antennas*.

If the spacing d between the array elements becomes greater than half a wavelength λ_0 , the denominator of $f_0(\psi)$ can have further zeros, resulting in additional array radiation beams. These beams are referred to as grating lobes. In practice, suppressing such parasitic lobes is required.

Figure 1-21 illustrates several array patterns derived from $f_0(\psi)$. These show the characteristic narrow main beam and a larger number of sidelobes as the number of array

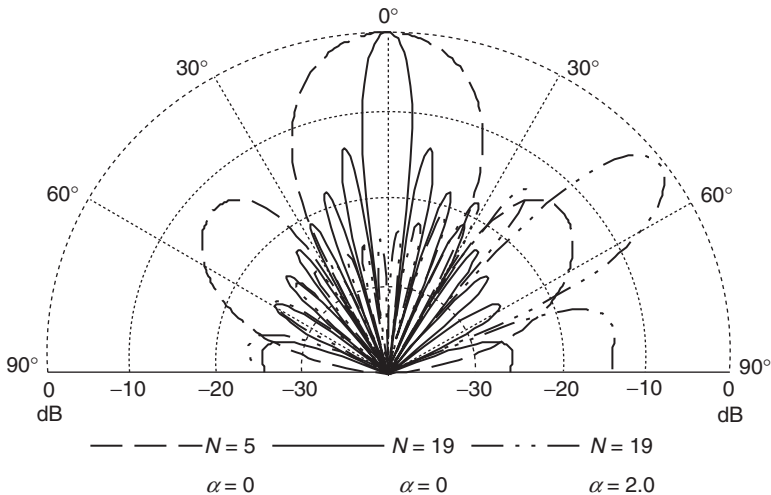


FIGURE 1-21 Linear array pattern factors f_0 (array element spacing $d = 0.4\lambda_0$)

elements increases. Lower sidelobe levels can be achieved by tapering the excitation of the array elements. Several techniques, such as the binomial or the popular Chebyshev methods (see Chapter 3 on array design) are used for controlling the sidelobe levels and the main beamwidth.

Next, we consider the 2D linear array in Figure 1-20b. This array has M elements along the z -axis and N elements along the y -axis. The array elements are equally spaced with corresponding spacings of d_z and d_y and associated linear phase progressions of α_z and α_y . For the computation of the array factor $f(\theta, \psi)$, the 2D array can be viewed as M equally spaced linear arrays along the z -direction. It is given by

$$\begin{aligned} f(\theta, \phi) &= \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} a_{mn} e^{jn(k_0 d_z \cos \theta - \alpha_z)} \\ &= \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} a_{mn} e^{jn\psi_z} e^{jm\psi_y} \end{aligned} \quad (1-42)$$

where $\psi_z = k_0 d_z \cos \theta - \alpha_z$ and $\psi_y = k_0 d_y \sin \theta - \alpha_y$. Assuming the amplitudes of all array elements to be identical ($a_{00} = a_0 = \dots = a_{M-1N-1}$) and applying the same manipulations as in the case of the linear array, the normalized array factor $f_0(\theta, \psi) = f_0(\theta, \psi)/(MNa_{00})$ is found to be

$$f_0(\theta, \phi) = \frac{\sin\left(\frac{Nk_0 d_z}{2} \cos \theta\right) \sin\left(\frac{Mk_0 d_y}{2} \sin \theta \sin \phi\right)}{N \sin\left(\frac{k_0 d_z}{2} \cos \theta\right) M \sin\left(\frac{k_0 d_y}{2} \sin \theta \sin \phi\right)} \quad (1-43)$$

In general, the amplitude of every array element can be chosen independently in order to shape arbitrary array patterns. Conventional beam-shaping techniques try to approximate directivity patterns known from continuous aperture distributions (Sections 1.7 and 1.8). More recently, digital signal processing techniques are employed to realize a large variety of applications such as angle of arrival detection (see Chapter 47 on direction finding), tracking, interferer suppression, and adaptive signal to interference improvement.

A severe disadvantage of linear array configurations is their restricted field of view. Reasonable beam-shaping and scanning can be achieved by modifying the array element amplitudes. However, many arrays have restricted angular range. This drawback can be overcome by array configurations *conforming* to curved surfaces. As such, every array element may have a different geometrical orientation, and suitable element amplitude directivity behavior can be a difficult task. Beam-shaping techniques have been developed incorporating digital signal-processing techniques for the necessary flexibility to achieve desirable designs (see Chapter 22).

A disadvantage of an array conformed to a curved surface is that not all antenna elements contribute to the radiated fields in a particular direction. Thus, more array elements are required to achieve certain requirements with respect to beamwidths and sidelobe suppression. An example of a singly curved conformal array antenna is shown in Figure 1-22.

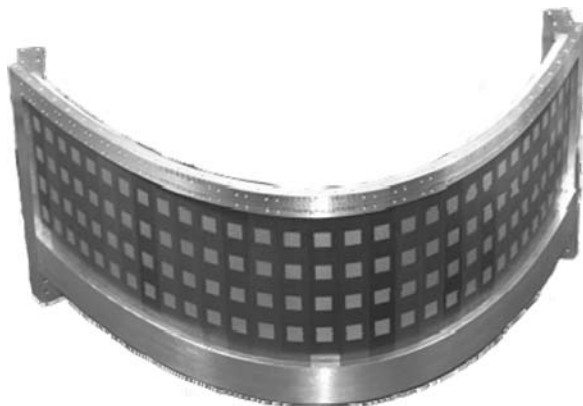


FIGURE 1-22 Singly curved array antenna realized with patch antenna elements (Courtesy of FGAN e.V., Wachtberg, Germany)

Another array configuration are *volumetric arrays*. To avoid shadowing of the individual array elements, the grid must be relatively thin, and reasonable beam-shaping is achieved only by randomly distributing array elements to suppress grating lobes due to large element spacing. In principal, all array elements can radiate in all directions. However, practical realization of volumetric arrays is not easy, and the choice of suitable array elements and feeding techniques is restricted. An example of a volumetric array is depicted in Figure 1-23.



FIGURE 1-23 Volumetric array antenna with randomly distributed loop elements: Crow's nest antenna by FGAN e.V., Wachtberg, Germany²⁰ (Courtesy of FGAN e.V., Wachtberg, Germany)

1.10 BASIC CONCEPTS IN MOBILE COMMUNICATIONS

Mobile communications is an area where unguided electromagnetic waves and antennas play an important role. The goal is to offer communication links between any place on earth without limiting user mobility. Given the restricted spectrum availability, the need for frequency reuse, particularly in densely populated areas, creates special challenges that are most efficiently resolved within terrestrial mobile communication systems. However, the requirement to cover remote regions makes essential the use of satellite systems in the overall strategy.

Terrestrial mobile communication systems are usually based on cellular principles, where the mobile terminal communicates with a fixed base station. Alternative strategies can, for instance, establish a connection via a network of mobile terminals only (*ad-hoc networks*); however, the focus here will be on cellular techniques. The fundamental issue in mobile communications is the restricted availability of frequency bands. Therefore, system design must aim at a high spectrum efficiency expressed in Erlang per square meter per Hertz and given by

$$\eta_s = \frac{\text{number of reuses}}{\text{coverage area}} \times \frac{\text{number of channels}}{\text{bandwidth}} \times \frac{\text{time the channel is busy}}{\text{total time of channel}} \quad (1-44)$$

where one busy communication channel is equivalent to the traffic of one Erlang.

The basic behavior of cellular designs is often studied in the context of hexagonal cell coverage, as illustrated in Figure 1-24, where the assumption is that the base station is located at the center of a cell. The idea is to reuse a given set of communication channels or frequencies (in those cells) that are sufficiently apart from each other so that co-channel interference remains within acceptable limits. Code division multiple access (CDMA) techniques assign relatively broad frequency bands to individual cells. Those techniques, however, can often work with the same frequency in neighboring cells since channel separation is achieved on a code level and frequency planning may be replaced or supplemented by code planning. Therefore, discuss reuse considerations on the channel level. The group of cells not employing channel reuse is called a *cluster*. Such a cluster is depicted in Figure 1-24 as a grey-shaded area and consists of seven cells.

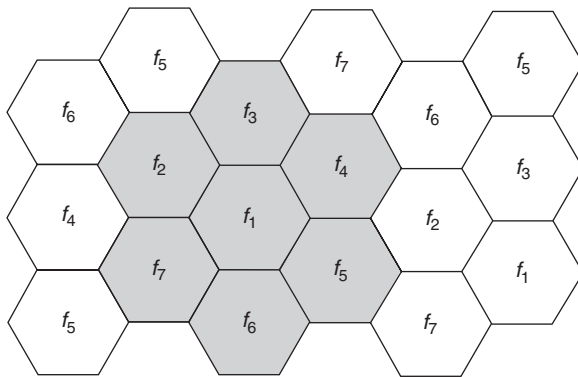


FIGURE 1-24 Hexagonal cellular pattern with a reuse factor of 7

The number of cells N comprising a cluster is, in general, given by

$$N = i^2 + ij + j^2 \quad (1-45)$$

where i and j are integers, with $N=7$ being a popular number of cells per cluster. Obviously, you can increase the mobile system's capacity (i.e., cellular phone users) by increasing the number of available channels (bandwidth) or the spectrum efficiency η_s . Because bandwidth is restricted, the logical approach is to achieve better spectrum efficiency through channel reuse. This implies smaller cell sizes and consequently a greater number of cells in a given coverage area. For example, a single cell can be subdivided into several smaller cells, or a cell can be divided into sectors by using directional antennas to avoid installing additional base stations. Typical cell dimensions used to cover wide rural regions or suburban areas are referred to as *macrocells*. In densely populated urban regions, however, cells must often be installed at the street level with base stations located below building roofs. Such cells are often called *microcells*, and even smaller cells (restricted to a single site or the interior of a building) are referred to as *picocells*.

The key to designing complex cellular mobile communication systems is electromagnetic wave propagation within individual cells. Propagation models are essential for evaluating modulation and coding schemes and their associated signal power within the service area or the interference that may be caused outside the service area. Even when adaptive power control is used at the transmitter, a thorough understanding of the relevant wave propagation mechanisms in the terrestrial environment is necessary. For propagation above a smooth ground, a simple ray-optical model including two rays is sufficient for modeling wave propagation. In general, ray approaches are based on high-frequency assumptions, typically fulfilled at mobile communication frequencies (greater than 400 MHz).

The two-ray model includes a direct ray representing free-space propagation and a ray reflected at the earth's surface. Assuming the earth's surface to be planar, the two-ray model gives the following pathloss:

$$L_p = 20 \log \left(\frac{\lambda}{4\pi r} \right) + 20 \log \left| 1 + R_b \frac{r}{r_b} e^{-j2\pi/\lambda (r_b - r)} \right| \quad (1-46)$$

L_p is the ratio of received to transmitted power in dB, where both antennas are assumed to be isotropic. The involved parameters are λ , or the free-space wavelength; r , or the path length of the direct ray; r_b , or the path length of the reflected ray; R_b , or the plane wave reflection coefficient at the reflection point.

Figure 1-25 compares L_p as given to single/direct-ray (free-space) and four-ray models for different transmitter antenna heights. The typical two-ray interference pattern is clearly identified close to the transmitter. Also, at great distances from the transmitter, the two-ray model predicts a pathloss that has a constant slope of 40 dB/decade versus a pathloss with a slope of 20 dB/decade predicted by the single-ray or free-space model. Figure 1-25 also shows that changing the transmitter antenna height can control the coverage range of the base station.

Microcells or picocells are often designed at the street level, and for these cases, the two-ray model for flat earth can be improved by including ray reflections from the street side walls. For example, Figure 1-25 shows four-ray pathloss curves to model a 16 m-wide street canyon, where contributions from the two sidewall-reflected rays are added to the two-ray results. Again, the usual interference patterns are observed close to the transmitter. However, at great distances from the transmitter, the guiding effect of the street canyon results in a pathloss that is even less than that predicted by the single-ray or free-space model. Also, the height of the transmitter antenna no longer influences the pathloss at great distances.

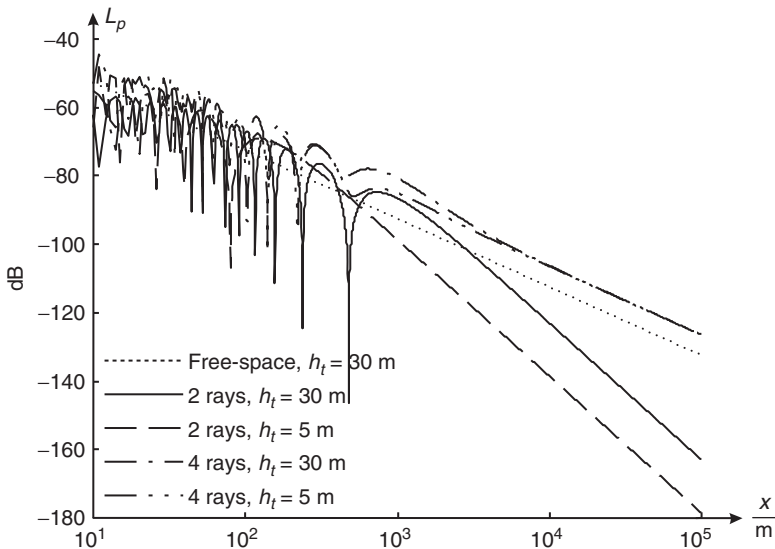


FIGURE 1-25 Pathlosses for different ray models over flat ground. Scalar computations with a receiver antenna height of 2.4 m, different transmitter antenna heights h_t , and all reflection coefficients set to -1 . The given four-ray models also include sidewall-reflected rays at each side of a 16 m broad street canyon.

Previously, we discussed simple propagation models. However, real-world terrestrial mobile communication channels are characterized by complicated wave propagation mechanisms. The received signal is composed of an extremely large number of different multiply reflected, diffracted, or scattered wave contributions, and a purely deterministic description of the radio channel might be impossible. Therefore, most receiver power or receiver field strength determination methods are designed to predict the corresponding median values as a first characterization of the channel properties. The fading behavior of the mobile radio channel is described by means of statistical methods where fast- and slow-fading mechanisms are distinguished. Fast-fading can often be characterized by Rayleigh or Rice probability distributions, and slow-fading normally behaves as lognormal distributed with standard deviations of several dBs.

Further insight into the behavior of a mobile radio channel can be gained from measured or predicted impulse responses, as shown in Figure 1-26, which typically consist of various signal contributions arriving at the receiver after different delay times and with different Doppler shifts due to moving transmitter and receiver antennas as well as scattering objects. Information that can be obtained from the impulse responses are the relative signal powers and delay spreads for the different signal contributions. Both the Doppler shifts and the delay spreads are essential in the design of mobile communication systems. Often, test sequences are included in the transmitted signals, which allow for the estimation of important channel parameters and can be used for channel equalization.

Modern terrestrial communication systems often utilize several transmit and/or receive antennas (*multiple input multiple output (MIMO)*) to achieve improved signal to noise and/or signal to interference ratios of the communication links. The different antennas can be arranged in closely spaced array configurations suitable for the realization of deterministic beam-forming strategies (see Section 1.8). Another strategy tries to arrange the antennas

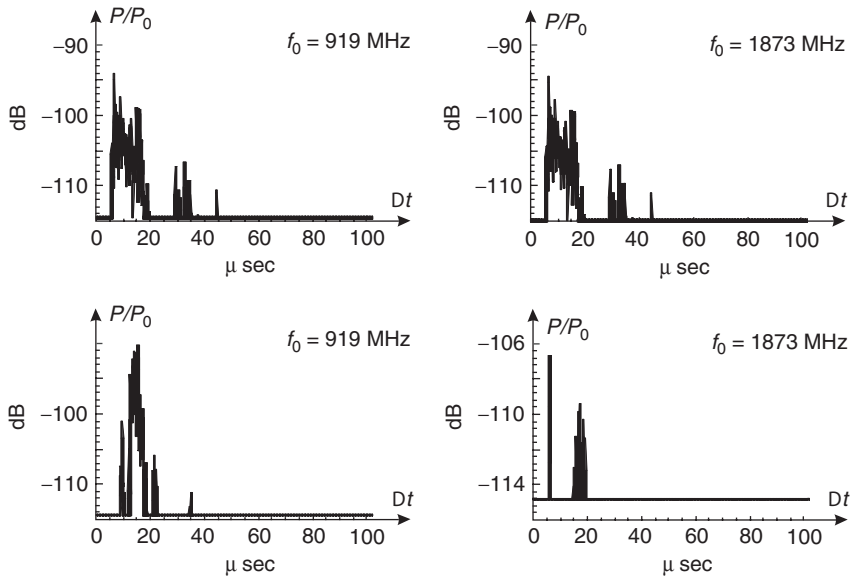


FIGURE 1-26 Measured impulse response of different mobile radio channels in Garmisch Partenkirchen, Germany (hilly terrain). The measurement bandwidth was 6 MHz at the given carrier frequency f_0 . The measured data shown were provided by Deutsche Telekom, Darmstadt, Germany.

such that their transmitting or receiving characteristics become uncorrelated, leading to so-called diversity techniques. Further information on terrestrial mobile communication systems can be found in Chapter 22 on conformal antennas.

REFERENCES

1. C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd Ed. (New York: John Wiley & Sons, Inc., 1996).
2. J. D. Kraus, *Antennas*, 2nd Ed. (New York: McGraw-Hill, 1988).
3. J. A. Kong, *Electromagnetic Wave Theory* (New York: John Wiley & Sons, Inc., 1990).
4. D. Pozar and D. Schaubert, *Microstrip Antennas* (Piscataway: IEEE Press, 1995).
5. S. Silver, *Microwave Antenna Theory and Design* (New York: McGraw-Hill, 1949): sec. 2.14.
6. H. T. Friis, "A Note on a Simple Transmission Formula," *IRE Proc.* (May 1946): 254–256.
7. *IEEE Standard Test Procedures for Antennas*, IEEE Std. 149-1979 (New York: Institute of Electrical and Electronics Engineers, 1979): sec. 11.
8. J. S. Hollis et al., *Techniques of Microwave Antenna Measurements* (New York: John Wiley & Sons, Inc., 1984).
9. R. C. Spencer and P. M. Austin, "Tables and Methods of Calculation for Line Sources," MIT Rad. Lab Rep. 762-2 (March 1946); see also Rep. 762-1.
10. J. F. Ramsay, "Fourier Transform in Aerial Theory," *Marconi Rev.*, vol. 9 (1946): 139; vol. 10 (1947): 17, 41, 81, 157.
11. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd Ed. (New York: John Wiley & Sons, Inc., 1998).

12. R. W. P. King, *The Theory of Linear Antennas* (Cambridge: Harvard University Press, 1956).
13. E. Jahnke and F. Emde, *Tables of Functions* (New York: Dover Publications, Inc., 1943): 227.
14. A. W. Love, *Electromagnetic Horn Antennas* (New York: IEEE Press, 1976).
15. P. J. B. Clarricoats and G. T. Poulton, "High-Efficiency Microwave Reflector Antennas—A Review," *Proc. IEEE* (1977): 1470–1502.
16. Y. T. Lo and S. W. Lee, *Antenna Handbook*, Volumes I–III. (New York: Van Nostrand Reinhold, 1993).
17. A. W. Rudge, "Offset-Parabolic Reflector Antennas: A Review," *Proc. IEEE* 66 (1943): 1592–1618.
18. J. B. Pendry, "Negative Refraction Makes a Perfect Lens," *Phys. Rev. Lett.* vol. 85 (October 2000): 3966–3969.
19. J. Brown, "The Effect of a Periodic Variation in the Field Intensity Across a Radiating Aperture," *IEE Proc (London)* part III, vol. 97 (November 1950): 419–424.
20. J. Ender and H. Wilden, "The Crow's Nest Antenna—a Spatial Array in Theory and Experiment," *Intern. Conf. on Antennas and Propagation* (1981): 25–27.

Chapter 2

Frequency Bands for Military and Commercial Applications

Derek M. K. Ah Yo

Oceanit

Rudy Emrick

Motorola

CONTENTS

2.1	INTRODUCTION TO FREQUENCY BANDS	2-2
2.2	FREQUENCIES AND TECHNOLOGIES OF INTEREST FOR MILITARY APPLICATIONS	2-6
2.3	FREQUENCIES AND TECHNOLOGIES OF INTEREST FOR COMMERCIAL APPLICATIONS	2-9
2.4	EXAMPLES OF FUNDAMENTAL ANTENNA TYPES USED IN MILITARY AND COMMERCIAL APPLICATIONS	2-15

2.1 INTRODUCTION TO FREQUENCY BANDS

The spectrum chosen for use in either military or commercial applications often depends on a number of factors, including regulatory requirements or licenses that designate bands for certain uses. Depending on the application, the frequency band selected may also depend on antenna size, propagation distance, or environment. Over time, a number of different designations for frequency bands have been developed. For comparison, each of these band designations is shown in Table 2-1. Historically, some of the band groupings have differed, in part based on the application(s) intended, such as radar, electric countermeasures, and so on. The most commonly used designations are also listed in Table 2-1.

As mentioned, one of the factors to consider when selecting a band is the propagation characteristics of the band. Figure 2-1 shows attenuation as a function of wavelength or frequency. Attenuation is also a function of a number of factors that include both scattering losses and absorption losses. In general, lower frequencies tend to propagate farther and to transmit better under non-line-of-sight conditions. However, lower frequency antennas also tend to be larger because antenna size scales with frequency, creating a number of trade-offs that must be considered. For example, the military might select a lower frequency for operation so communication over mountain ranges is optimized; however, the lower frequency creates challenges in implementing an antenna that is a

TABLE 2-1 Frequency-Band Designations

IEEE Radar Bands*		ITU Frequency Bands [†]		Common-usage Bands [‡]		Electric-countermeasure Bands [§]	
Band	Frequency Range (in GHz)	Band	Frequency Range (in GHz)	Band	Frequency Range (in GHz)	Band	Frequency Range (in GHz)
HF	0.003–0.03	HF	0.003–0.03	HF	0.003–0.03	A	0–0.25
VHF	0.03–0.3	VHF	0.03–0.3	VHF	0.03–0.3	B	0.25–0.5
UHF	0.3–1	UHF	0.3–3	UHF	0.3–1	C	0.5–1
L	1–2	SHF	3–30	L	1–2	D	1–2
S	2–4	EHF	30–300	S	2–4	E	2–3
C	4–8			C	4–8	F	3–4
X	8–12			X	8–12.4	G	4–6
Ku	12–18			Ku	12.4–18	H	6–8
K	18–27			K	18–26.5	I	8–10
Ka	27–40			Ka	26.5–40	J	10–20
mm	40–300			Q	33–50	K	20–40
				V	50–75	L	40–60
				W	75–110	M	60–100

*From Institute of Electrical and Electronic Engineers Standard 521-1976, Nov. 30, 1976.

[†]From International Telecommunications Union, Art. 2, Sec. 11, Geneva, 1959.

[‡]No official international standing.

[§]From AFR 55-44 (AR 105–86, OPNAVIST 3430.9B, MCO 3430.1), Oct. 27, 1964.

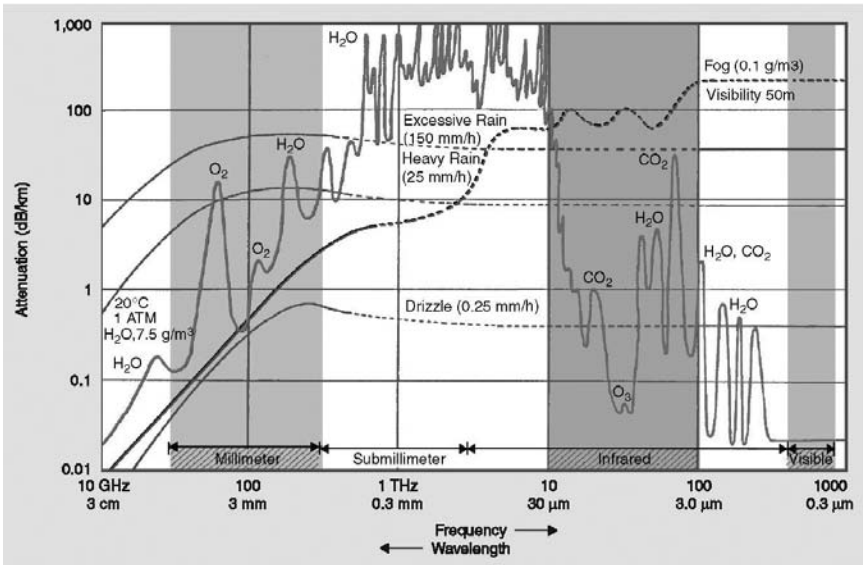


FIGURE 2-1 Attenuation of electromagnetic waves as a function of frequency or wavelength (after L. Yujiri et al¹ © IEEE 2003)

reasonable size. In other cases, the military might select a band where propagation is limited for security or frequency re-use opportunities. An example of such a band is the V-band around 60 GHz. This band is well suited to short-range applications, but the implementation of low-cost antennas that perform suitably can be difficult due to the high frequency of operation, though the antennas will be relatively small for a given antenna gain compared to the lower frequencies.

The trade-offs associated with the use of various frequency bands have been studied extensively over time, and certain characteristics and applications are now commonly linked to specific frequency bands. Table 2-2 lists a number of frequency bands along with their common characteristics and uses. In general, the lower frequency bands tend to be used for longer range, lower bandwidth applications while higher bands tend to be used more often for shorter range applications that require higher bandwidth. In the case of radar, the higher frequency bands are often selected for higher resolution and the antennas take up less volume at these higher frequencies.

In the case of television and radio broadcasts, the selected bands tend to be lower, such as VHF and UHF, because the broadcasting company will likely want to cover an entire city or an even greater area with a single transmit site. The specific frequency band used by a particular station also depends on its license through a regulatory agency such as the Federal Communications Commission (FCC), which designates specific channels or bands with well-defined separation between them to avoid interference with broadcast services. The bands and channels used for AM/FM radio and television broadcast in the U.S. are shown in Table 2-3. The separation between the bands, sometimes called *guard bands*, help prevent interference between adjacent bands, but these bands were established decades ago. With continuing advances in communications technology and equipment, and the move toward digital television, the FCC is looking at different ways of thinking about the bandwidth per channel and the guard bands. In transmitting digital television, less bandwidth is required than

TABLE 2-2 Common Frequency Bands: Characteristics and Typical Applications

Band	Characteristics	Applications
HF	Long-distance links possible with ionosphere reflections	Ocean vessel communications, telephone, telegraphy, long-range aeronautical communications, amateur radio communications, military communications
VHF	Ionosphere reflections possible at the lower end of the frequency band	Television and FM broadcasting, air traffic control, radio navigation, military communications
UHF	Tends to require more line-of-sight toward the high end of the frequency band	Television broadcasting, radar, mobile phones and radios, satellite communications, global positioning systems (GPSs), wireless LANs, wireless PANs, military communications
SHF	Atmospheric absorption at highest frequencies can be significant	Radar, microwave links, land mobile communications, satellite communications, direct broadcast satellite (DBS) television
EHF	Line-of-sight propagation only toward the high end of the band, subject to atmospheric absorption, and best suited to shorter range applications	Radar, secure and military communications, satellite links, gigabit per second backhaul (1–2 km), future wireless PANs
Optical / IR	Significant atmospheric absorption, typically will not penetrate fog, line-of-sight only	Optical communications, fiber-optical links, very short range wireless communications, building-to-building high-speed wireless connections

with analog, but the bandwidth does depend on the resolution or definition of the signal. In the future, digital television may evolve toward either one very high-definition digital signal or multiple lower definition signals for multiple television channels in what was one signal when transmitting analog. Advances in communications technology have also led to the discussion of the potential use of the guard bands for other applications since advances may allow for their use without interfering with the original broadcast channels.

TABLE 2-3 Typical Broadcast Frequencies

VHF TELEVISION FREQUENCIES					
Band	Ch #	Frequency	Band	Ch #	Frequency
VHF LOW	02	54–60 MHz	VHF HIGH	07	174–180 MHz
VHF LOW	03	60–66 MHz	VHF HIGH	08	180–186 MHz
VHF LOW	04	66–72 MHz	VHF HIGH	09	186–192 MHz
VHF LOW	05	76–82 MHz	VHF HIGH	10	192–198 MHz
VHF LOW	06	82–88 MHz	VHF HIGH	11	198–204 MHz
			VHF HIGH	12	204–210 MHz
			VHF HIGH	13	210–216 MHz

TABLE 2-3 (continued)

UHF TELEVISION FREQUENCIES					
Ch #	Frequency	Ch #	Frequency	Ch #	Frequency
14	470–476 MHz	38	614–620 MHz	62	758–764 MHz
15	476–482 MHz	39	620–626 MHz	63	764–770 MHz
16	482–488 MHz	40	626–632 MHz	64	770–776 MHz
17	488–494 MHz	41	632–638 MHz	65	776–782 MHz
18	494–500 MHz	42	638–644 MHz	66	782–788 MHz
19	500–506 MHz	43	644–650 MHz	67	788–794 MHz
20	506–512 MHz	44	650–656 MHz	68	794–800 MHz
21	512–518 MHz	45	656–662 MHz	69	800–806 MHz
22	518–524 MHz	46	662–668 MHz	70	806–812 MHz
23	524–530 MHz	47	668–674 MHz	71	812–818 MHz
24	530–536 MHz	48	674–680 MHz	72	818–824 MHz
25	536–542 MHz	49	680–686 MHz	73	824–830 MHz
26	542–548 MHz	50	686–692 MHz	74	830–836 MHz
27	548–554 MHz	51	692–698 MHz	75	836–842 MHz
28	554–560 MHz	52	698–704 MHz	76	842–848 MHz
29	560–566 MHz	53	704–710 MHz	77	848–854 MHz
30	566–572 MHz	54	710–716 MHz	78	854–860 MHz
31	572–578 MHz	55	716–722 MHz	79	860–866 MHz
32	578–584 MHz	56	722–728 MHz	80	866–872 MHz
33	584–590 MHz	57	728–734 MHz	81	872–878 MHz
34	590–596 MHz	58	734–740 MHz	82	878–884 MHz
35	596–602 MHz	59	740–746 MHz	83	884–890 MHz
36	602–608 MHz	60	746–752 MHz		
37	608–614 MHz	61	752–758 MHz		

Broadcast Frequencies

AM Radio = 535 kHz–1605 kHz (MF) 107 Channels each with 10 KHz separation

TV Band I (Channels 2–6) = 54 MHz–88 MHz (VHF)

FM Radio Band II = 88 MHz–108 MHz (VHF) 100 Channels each with 200 KHz separation

TV Band III (Channels 7–13) = 174 MHz–216 MHz (VHF)

TV Bands IV & V (Channels 14–69) = 470 MHz–806 MHz (UHF)

In addition to commercial broadcast channels, a number of amateur radio bands have been established. These bands also tend to be at lower frequencies since it is usually desirable to have propagation over long distances to communicate with other amateur radio sites, as shown in Table 2-4. The successful use of the higher bands depends on the proximity of other amateur sites because they tend to propagate over shorter distances.

TABLE 2-4 Amateur Radio Bands

Band	Frequency
160 m	1.8–2.0 MHz
80 m	3.5–4.0 MHz
40 m	7.0–7.3 MHz
20 m	14.0–14.35 MHz
15 m	21.0–21.45 MHz
10 m	28.0–29.7 MHz
6 m	50.0–54.0 MHz
2 m	144.0–148.0 MHz
	220–225 MHz
	420–450 MHz
	1215–1300 MHz
	2300–2450 MHz
	3300–3500 MHz
	5650–5925 MHz

2.2 FREQUENCIES AND TECHNOLOGIES OF INTEREST FOR MILITARY APPLICATIONS

Military and commercial communication needs have similarities and differences, as shown in Figure 2-2 and Figure 2-4, later in the chapter in Section 2.3. One area of commonality is that both stem from the demand for quick and efficient delivery of information. However, the pace at which they develop their communication technologies and their dynamic motivations for doing so are where the two sectors diverge and are continuously evolving. In some cases, the commercial sector’s communication development outpaces the military’s because its driving force is economic and its market is vast. In other instances, military technology outpaces the commercial sector since performance may be more important than low cost for critical equipment. In some ways, the two sectors are merging with the military’s technology, research, and development organizations such as Defense Advanced Research Projects Agency (DARPA), Office of Naval Research (ONR), Naval Research Laboratory (NRL), Air Force Research Laboratory (AFRL), and Space and Naval Warfare (SPAWAR) Systems Command, moving toward leveraging more commercial technology as the military increases its demand for commercial sector innovations to bring its communication systems up to 21st-century standards. The military can optimize its communication systems by complementing available commercial technology with newly developed, customized technology providing a superior hybrid solution.

Most terrestrial communication systems used by the military and the commercial market, including HF, VHF, and UHF radios, utilize conventional monopole and dipole antennas that are electrically small but long enough to provide adequate gain and the largest possible coverage area. Most *satellite communication (satcom)* systems depend on mechanically rotated reflector dish antennas that provide high gain and a narrow beam from a relatively simple parabolic antenna structure in order to illuminate the orbiting satellites with precision accuracy. These antenna structures are, however, still large and can be used as a visual cue in the identification of the communication system’s location. In the case of the military,

		Wireless Technology	Frequency Band	Frequency	Free Space λ	Range	Data Rate	Deploy Date	Comm Devices/ Operation	Antenna Technologies	
										Traditional	Compact
Military	Terrestrial Comm	PRC-150	HF	2-30 MHz	10-150 m	30+ miles	9.6-14.4 kbps	2001	secure voice/data/ networking	Mono-Dipole	Spherical Helix
		RT-1523	VHF	30-88 MHz	3.4-10 m	10-100s miles	9.6-14.4 kbps	1994	SINGARS single ch gnd&air radio system	Mono-Dipole	Spherical Helix
		PRC-148	VHF-UHF	30-512 MHz	0.5-10 m	12 miles	NA	2001	Voice/low rate data	Mono-Dipole	Spherical Helix
		PRC-117	VHF-UHF	30-512 MHz	0.5-10 m	10-50 miles	NA	NA	secure voice/data LOS & UHF Satcom	Mono-Dipole	Spherical Helix
		PSC-5D	VHF-UHF	30-512 MHz	0.5-10 m	10-50 miles	76.8 kbps	1997	secure voice/data LOS & UHF Satcom	Mono-Dipole	Spherical Helix
		RT-1720 EPLRS	UHF	20-450 MHz	66-71 cm	6-60 miles	486 kbps	2000	secure voice/data/ networking	Mono-Dipole	PIFA
		VRC-99	L-band	1308-1484 MHz	15-23 cm	150 miles w/ relay	625 kbps; 10 Mbps bursts	NA	Voice/data/ video/ network radio	Mono-Dipole	Half-Disc Ant
		SecNet11 Secure WLAN	S-band	2.412-2.462 GHz	12.5 cm	120 m	1-11 Mbps	NA	Secure Wireless LAN	Mono-Dipole	Patch
		UHF TacSat	UHF	243-318 MHz	.94-1.23 m	Earth to LEO	NA	NA	UHF Tac Sat Comm	Dish Ant	Phased Array
		Ku-Satcom	Ku-band	Uplink: 11.2-11.7 Ghz downlink: 14-14.5 GHz	2-2.7 cm	Earth to GEO	.5-5 Mbps	Late 1970s	Ku-band Sat Comm System	Mech Dish Ant	Electronic Scan Phased Array
	Satellite Comm	Ka-Satcom	Ka-band	Uplink: 27.5, 31 GHz downlink: 18.3, 18.8, 19.7, 20.2 GHz	1-1.6 cm	Earth to GEO	Uplink 2 Mbps downlink 30 Mbps	April 2005	Ka-band Sat Comm System	Mech Dish Ant	Electronic Scan Phased Array

FIGURE 2-2 Wireless technologies for military communication systems

inconspicuous communication vehicles are essential to protecting communication capability and the personnel both operating the system and receiving its intelligence support. The commercial sector also benefits from reduced antenna architectures because the savings in real estate allow for additional communication systems to be installed. In the case of cellular infrastructure, reduced size systems can translate into reduced recurring costs because less space may need to be leased at the site's location.

Methods for reducing antenna size include using planar antenna geometries similar to those of microstrip antennas with parasitic shorting structures,²⁻⁵ high dielectric substrates,⁶ geometrical variations of the antenna architecture including slots,⁷⁻¹⁰ and active element integration into the antenna.

Connectivity on today's battlefield is of such great importance that communication development programs are being forced to develop rapidly. Multifunction communication platforms are being built to satisfy the need for multiple radio access and centralized intelligence distribution using a router methodology that consolidates the information from several different radio sources and redistributes it to personnel using the appropriate communication system. A multitude of processing and communication systems are utilized together to keep all mission participants apprised of tactical and strategic information as it becomes available, allowing command decisions to be handed down as required.

The radios used by the military vary in frequency and constitute a range of frequencies that span from the HF to Ka-band, as shown in Figure 2-3.

Tactical radios in the HF frequency, including the PRC-150 from 2 MHz to 30 MHz, are important in transmitting voice and limited data over long distances by pumping large amounts of power (20–100 W) through substantially large monopole antennas. These tactical radios also fall within the civilian Ham radio frequencies, which also require large amounts of power to operate.

VHF radios operate at slightly higher frequencies from 30–88 MHz, providing for increased bandwidth to transmit more data over the same long distances. This is true for the SINCGARS RT-1523E radio system, which also pumps large amounts of power through a large monopole antenna to extend the range of its transmissions.

Radios that operate at UHF, which also include the L-band and S-band from 300 MHz to 3 GHz, provide increased bandwidth capability and faster data rates than HF and VHF systems but at the expense of reduced transmission range. Operating at these higher frequencies also scales down the size of the antennas used to Tx/Rx signals, which reduces the visibility of mobile communication platforms. As presented in Figure 2-2, the PRC-148, PRC-117, PSC-5D, RT-1720 EPLRS, and VRC-99 radios all operate at these frequencies. The range of the systems are reduced due to the limitations of shorter

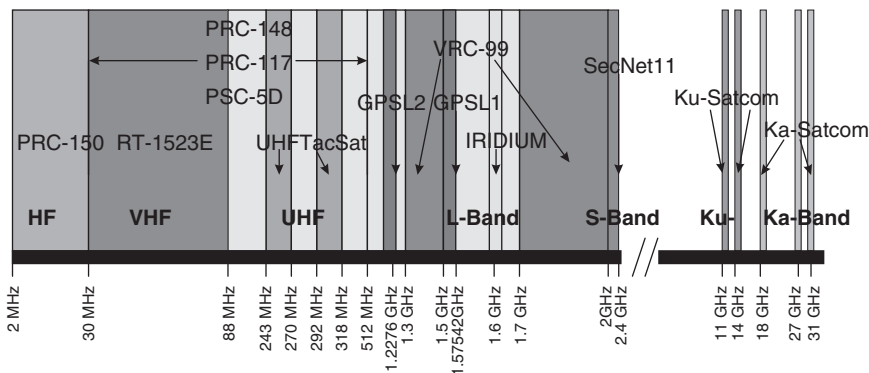


FIGURE 2-3 Military communications frequency spectrum

wavelength transmissions, but this shortfall is more than compensated for because larger bandwidths are available and long-distance data transfers can be achieved by using relays to leapfrog signals between intermediate points and then on to remote positions.

Situational awareness reports that include highly detailed maps with reconnaissance pictures and even video are transmitted by tactical satellite radios operating at the C-, Ku-, and Ka-band. These frequencies are capable of penetrating through the ionosphere to orbiting satellites and then back down to other positions on earth that are over-the-horizon from the position of the originating source transmission. This over-the-horizon capability and increased bandwidth are major goals of the military and also of the commercial market, which means greater amounts of data can be sent more quickly to more distant locations thus providing needed intelligence for informed decision making.

This wide array of radio systems over the large frequency range increases military capability, especially when placed on a mobile platform, but at the cost of increased hardware including the radios themselves, networking equipment, power sources, and the antennas to perform the necessary signal propagation. This excessive amount of hardware is expensive and bulky, and in the case of the antennas, presents the problem of physically placing them on top of a small roof, deck, mast, or fuselage. Not only are there co-site signal propagation issues associated with multiple antennas for multiple communication systems, but also there is the burden of concealing the vehicle, ship, or aircraft's purpose from unwanted spectators. In order to decrease the visibility of the mobile platform as a target and maximize its carrying capacity, there is a need to scale down the size, weight, and volume of all components including the antennas. This requires antenna designs that use planar antenna technologies and allow for dual-band, multiband, and broadband multifunction capability.

2.3 FREQUENCIES AND TECHNOLOGIES OF INTEREST FOR COMMERCIAL APPLICATIONS

Commercial frequency bands and application technologies are driven by a number of factors but none more significant than cost. As previously discussed, this factor is more relevant for commercial applications than for military ones. The overall cost of a solution or service is a direct function of equipment cost but can also be affected by the spectrum selected for use. In some cases, licensing spectrums can be very expensive and that cost must eventually be passed on to the consumer by a service provider. One of the advantages of using the licensed spectrum is that there is recourse if another company or individual interferes with the spectrum. The unlicensed spectrum is typically open for use by anyone as long as the conditions and limits set by that country's regulatory agency, such as the FCC in the U.S., are adhered to, but there is no recourse if another user creates interference when using the spectrum in the same vicinity.

The backbone of communication is the allocated frequency bands that radio equipment use to transmit and receive its data, and there are various wireless networks and protocols used for management. Commercial wireless protocols are varied and continue to evolve from the present second-generation (2G) to third-generation (3G) and future fourth-generation (4G) network system protocols that provide expanded bandwidth and increased data-rate capability. The 2G protocols were based on low-band digital data signals that provided a vast improvement over the previous analog systems. The 3G protocols allow wideband digital data transfers for large video and data transmissions. The trend is toward multimedia services, and this trend has made possible podcasting and streaming video that were once tethered to Ethernet cable connections on desktop computers but are now accessible with wireless multimedia phones, personal digital assistants (PDA), and laptops.

		Wireless Technology	Frequency Band	Free Space λ	Range	Data Rate	Deploy Date	Comm. Devices/Operation	Antenna Technologies	
									Traditional	Compact
Commercial	Digital Wireless Protocols	(2G) TDMA IS-54/IS-136 Time Division Multiple Access FDM	UHF - L-band	15-36 cm	100-10,000 m	48.6 Kbps	1994	TDMA cellular phones	Mono-Dipole	Patch Variant
		(2G) CDMA IS-95 Code Division Multiple Access FDM		14-36 cm	100-10,000 m	1.2288 Mbps	1996	Mobile phones on CDMA networks	Mono-Dipole	Patch Variant
		(2G) GSM Global System for Mobile Comm TDMA/FDM		15-34 cm	100-10,000 m	270.8 Kbps	1987 Europe 1995 USA	GSM enabled cell phones, PDAs, pagers	Mono-Dipole	Patch Variant
		(2G) DCS1800/ DCS1900 TDMA/FDM		16-18 cm	100-10,000 m	270.8 Kbps	mid 1990's	DCS enabled cell phones, PDAs, pagers	Mono-Dipole	Patch Variant
		(2G) PCS Personal Comm. System TDMA/FDM		20-37 cm	100-10,000 m	42 Kbps	1994	PCS enabled cell phones	Mono-Dipole	Patch Variant
		3G GSM 3G GSM network		14-16 cm	100-10,000 m	2 Mbps	?	3GSM-enabled cell phones, PDAs, pagers	Mono-Dipole	Patch Variant
		CDMA2000 3G CDMA based network		NA	100-10,000 m	144 Kbps future capability 4.8 Mbps	Feb 2003 Korea	Mobile phones on CDMA2000 networks	Mono-Dipole	Patch Variant
		(3G) GPRS General Packet Radio Service		NA	100-10,000 m	40-50 Kbps; 171 Kbps limit	2001	GPRS enabled cell phones/ networks Interface overlaid on existing GSM networks allowing for internet access	Mono-Dipole	Patch Variant
		AMPS/NAMPS Narrow Band Advanced Mobile Phone System FDMA		34-36 cm	100-10,000 m	NA	1978	AMPS/NAMPS enabled cellular phones	Mono-Dipole	Patch Variant
		NMT900 Nordic Mobile telephone FDMA		31-69 cm	100-10,000 m	NA	1986	NMT enabled cellular phones	Mono-Dipole	Patch Variant
Commercial	Analog Wireless Protocols	(2G) TDMA IS-54/IS-136 Time Division Multiple Access FDM	UHF - L-band	15-36 cm	100-10,000 m	48.6 Kbps	1994	TDMA cellular phones	Mono-Dipole	Patch Variant
		(2G) CDMA IS-95 Code Division Multiple Access FDM		14-36 cm	100-10,000 m	1.2288 Mbps	1996	Mobile phones on CDMA networks	Mono-Dipole	Patch Variant
		(2G) GSM Global System for Mobile Comm TDMA/FDM		15-34 cm	100-10,000 m	270.8 Kbps	1987 Europe 1995 USA	GSM enabled cell phones, PDAs, pagers	Mono-Dipole	Patch Variant
		(2G) DCS1800/ DCS1900 TDMA/FDM		16-18 cm	100-10,000 m	270.8 Kbps	mid 1990's	DCS enabled cell phones, PDAs, pagers	Mono-Dipole	Patch Variant
		(2G) PCS Personal Comm. System TDMA/FDM		20-37 cm	100-10,000 m	42 Kbps	1994	PCS enabled cell phones	Mono-Dipole	Patch Variant
		3G GSM 3G GSM network		14-16 cm	100-10,000 m	2 Mbps	?	3GSM-enabled cell phones, PDAs, pagers	Mono-Dipole	Patch Variant
		CDMA2000 3G CDMA based network		NA	100-10,000 m	144 Kbps future capability 4.8 Mbps	Feb 2003 Korea	Mobile phones on CDMA2000 networks	Mono-Dipole	Patch Variant
		(3G) GPRS General Packet Radio Service		NA	100-10,000 m	40-50 Kbps; 171 Kbps limit	2001	GPRS enabled cell phones/ networks Interface overlaid on existing GSM networks allowing for internet access	Mono-Dipole	Patch Variant
		AMPS/NAMPS Narrow Band Advanced Mobile Phone System FDMA		34-36 cm	100-10,000 m	NA	1978	AMPS/NAMPS enabled cellular phones	Mono-Dipole	Patch Variant
		NMT900 Nordic Mobile telephone FDMA		31-69 cm	100-10,000 m	NA	1986	NMT enabled cellular phones	Mono-Dipole	Patch Variant

FREQUENCY BANDS FOR MILITARY AND COMMERCIAL APPLICATIONS

2-11

		TACS	NTACS:	32–35 cm	100–10,000 m	NA	1988 UK	TACS enabled cellular phones	Mono-Dipole	Patch Variant
Commercial	TV Broadcast	Total Access Communication System	Rx: 860–870 MHz Tx: 915–925 MHz ETACS Rx: 916–949 MHz Tx: 871–904 MHz	32–35 cm	100–10,000 m	NA	1988 UK	TACS enabled cellular phones	Mono-Dipole	Patch Variant
		VHF TV	VHF	44–216 MHz	1.4–7 m	NA	NA	6 VHF Ch, 7 FM Ch	Yagi	NA
		UHFTV	UHF	470–806 MHz	37–64 cm	NA	NA	56 UHF Ch	Yagi	NA
		802.11a	C-band	5 GHz	6 cm	54 Mbps	1999	Wireless internet access to Laptop computers, PDAs, cell phones	Mono-Dipole	Patch Variant
		802.11b		2.4 GHz	12.5 cm	11 Mbps	1999	"	Mono-Dipole	Patch Variant
	Wireless LAN Protocols	802.11g		2.4 GHz	12.5 cm	54 Mbps	Jun-06	"	Mono-Dipole	Patch Variant
		802.11n	S-band	2.4 GHz	12.5 cm	540 Mbps	mid 2007	"	Mono-Dipole	Patch Variant
		802.15.1 Bluetooth		2.4 GHz	12.5 cm	720 Kbps	May-99	Printers, Cameras, cell peripherals	Mono-Dipole	Patch Variant
		802.15.4 ZigBee	ISM band Industrial scientific and medical	868 MHz, 915 MHz, 2.4 GHz	6 cm, 33 cm, 35 cm	100 Kbps	Jun-06	Bursts of power to extend battery for industrial data transfer, building and home automation	Mono-Dipole	Patch Variant
		(4G) 802.16 WiMax OFDM FDD/TDD	S-band C-band	2.5–2.69 GHz, 2.7–2.9 GHz, 3.4–3.6 GHz, 5.725–5.86 GHz	5–12 cm	70 Mbps	2004	Extended city/rural range wireless access w/ modem, PDA	Mono-Dipole	Patch Variant
Satellite Comm	Satellite Comm	(4G) Broadway HIPERLAN/2 HIPERSHOT OFDM	C-band W-band	5 GHz 59–65 GHz	0.5 cm 6 cm	100s Mbps 1–5 Gbps	2007?	Unlicensed band worldwide	Mono-Dipole	Patch Variant
		Iridium	L-band	1.616–1.628 GHz	18 cm	Earth to LEO	Sep-98	Satellite phone	Helical Ant	Circ polar microstrip
		C-band Satcom	C-band	Upink: 5.925–6.425 GHz downlink: 3.7–4.2 GHz	4.7–8.1 cm	Earth to GEO	1960s	C-band Sat Comm System	Mech Dish Ant	Electronic Scan Phased Array
		Ku-band Satcom	Ku-band	Upink: 11.2–11.7 GHz downlink: 14–14.5 GHz	2–2.7 cm	Earth to GEO	Late 1970s	Ku-band Sat Comm System	Mech Dish Ant	Electronic Scan Phased Array
		Ka-band Satcom	Ka-band	Upink: 27.5–31 GHz downlink: 18.3, 18.8, 19.7/20.2 GHz	1–1.6 cm	Earth to GEO	April 2005	Ka-band Sat Comm System	Mech Dish Ant	Electronic Scan Phased Array

FIGURE 2-4 Wireless technologies for commercial communication systems

The intermediate bridge between 2G and 3G is 2.5G, and it allows theoretically limited data transfer rates of up to 384 kbps, which is a significant increase from the previous 9.6–14.4 kbps capability of 2G. But, of course, theoretical limits are often just that—theoretical—when the reality is that 40–50 kbps is the nominal data rate.

In 2010–2015, 4G will succeed 3G and will offer increased data rates up to 100 Mbps while on the move and 1 Gbps while stationary for on-demand high-quality video transmissions and accelerated downloads. Several factors will contribute to 4G's succession, including the use of *Orthogonal Frequency Division Multiplexing (OFDM)* and *Orthogonal Frequency Division Multiple Access (OFDMA)* to more efficiently allocate resources between multiple users. Also, a network based solely on packet switching as opposed to the inefficient combined circuit-switched and packet-switched network is an improvement over 3G. Software-defined radio (SDR) receivers with wider bandwidth capability may also be used. In addition, 4G could be orchestrated using *pervasive networks* (a.k.a. ubiquitous computing) as first introduced by Mark Weiser,¹¹ which is a theoretical concept that allows multiple wireless technologies to be accessed simultaneously and information exchanged between them seamlessly while using 4G devices. To better manage the frequency spectrum allocation and transmission power utilized, pervasive networks will also use a *smart-radio* (a.k.a. cognitive radio) technology that decides the best frequency and power management.

Commercial wireless technologies include wireless cellular and wireless local area and personal area networks (WLAN, WPAN). WLAN systems typically operate in a range from 10–100 m while WPAN systems typically operate up to about 10 m. Cellular technologies have evolved from analog systems to today's higher capacity digital systems with increased speeds and capabilities. The system protocols deployed in the initial move to digital cellular, often referred to as 2G, varied greatly in method but all supported the common function of delivering information between users. The wireless protocols operate in the ~850/900/1800/1900/2100 MHz frequency bands, with the specific bands varying depending on the country, mode of operation, and cellular carrier. As shown in Figures 2-4 and 2-5, many digital protocols are in use today, such as time division multiple access (TDMA), code division multiple access (CDMA), Global System for Mobile Communications (GSM), digital communication systems (DCS), personal communication

Wireless Protocols	Transmission Rate (Mbps)	Effective Range (m)
802.15.3a UWB (disbanded)	50–500	1–10
802.11n	10–100	1–100
802.11a/g	5–50	1–100
802.16 WiMax	5–10	100–10,000
802.15.3	5–10	1–10
802.15.3c.MMV	100–5000	1–10
802.11b WiFi	1–10	1–100
802.16e/20	1	100–10,000
UMTS/HSDPA/txEVDO	0.5–1	50–10,000
802.15.1 Bluetooth	0.5–1	1–10
GPRS/EDGE	0.1–0.5	50–10,000
802.15.4 Zigbee	≤0.1	1–50
GSM/TDMA	≤0.1	50–10,000
Blackberry	≤0.1	50–50,000

FIGURE 2-5 Wireless protocols: their transmission rates (Mbps) and effective ranges (meters)

systems (PCS), and analog protocols that are still in use but in a diminishing capacity, for instance, Advanced Mobile Phone System (AMPS/NAMPS), Nordic Mobile Telephone (NMT), and Total Access Communication Systems (TACS). Worldwide today, GSM is the most widely used of these cellular protocols. Third-generation (3G) extensions of the digital protocols are emerging to increase bandwidth and data rates and include 3GSM, a 3G GSM network; CDMA2000, a 3G CDMA-based network; and General Packet Radio Services (GPRS). Some 3G protocols have been implemented in Asia already and are expected to be implemented in the U.S. and Europe in 2006–2007.

The 802.16 WiMAX protocol is one of the up-and-coming premier high-speed wireless broadband solutions that may provide greater than 70 Mbps data rates to high-density medium-size areas (1000–5000 m) such as city buildings, small and rural communities, and suburban malls. WiMAX operates at S-band and C-band frequencies and was first implemented into WiMAX indoor and outdoor modems in 2004 using orthogonal frequency division multiplexing. WiMAX will be further embedded into the wireless community's portable devices, including 3G and 4G ultra mobile PCs, over the next two to three years providing high-speed broadband wireless capability. However, WiMAX will still rely on an overlying wireless protocol, such as GSM or CDMA, to distribute the system over large city- and region-wide areas, causing some debate over how WiMAX will coexist with future cellular technologies if it is as widely deployed as expected.

WLAN and WPAN protocols (802.11 family) are also included in Figures 2-4 and 2-5 and are responsible for wireless data connections of various speeds and bandwidths to the internet for laptops, PDAs, and other wireless devices that operate in close proximity to one another. These include the 802.11 family operating at 2.4 GHz and 5 GHz. The 802.11 data rates continue to increase as well, with 802.11b offering 11 Mbps, 802.11a/b offering 54 Mbps, and the emerging 802.11n that's expected to offer data rates in excess of 100 Mbps, in part by utilizing multiple-input multiple-output (MIMO) communication techniques.

In addition, 802.15.4 ZigBee protocols, which allow devices to communicate low bandwidth and low data rate information between each other and achieve extremely long battery life, are expected to enable systems such as wireless sensor networks (WSNs) that can be deployed and left in place without maintenance for extended periods of time.

Bluetooth, which is covered under IEEE 802.15.1, is today's most widely deployed WPAN solution and is commonly used for wireless hands-free headsets for mobile phones. This area of wireless continues the trend toward higher speeds. There are some emerging technologies, however, that promise to move beyond the initial Bluetooth data rates of approximately 1 Mbps. Some of the potential solutions to offer very high speeds at short ranges include ultra wideband (UWB), dynamic frequency selection, cognitive radios, and higher-order modulation to achieve high throughputs in a limited spectrum with much higher spectral efficiency. Each of these potential solutions face technical challenges with some also confronting additional regulatory challenges. Dynamic frequency selection and cognitive radio, for example, may offer the shared use of large amounts of licensed spectrum, though it's not clear if regulatory bodies like the FCC and their equivalent in other countries will allow unlicensed users access to parts of the licensed spectrum or what the reaction will be from licensed users. UWB appears to be facing some regulatory challenges as Japan and Europe seem to have become more concerned with potential interference from using the full 3–10 GHz spectrum allowed in the U.S. and may limit its use to a subset of what the FCC has allowed. These bodies may limit wide scale deployment of UWB hardware to the resulting common worldwide spectrum, since a single solution for worldwide deployment is desirable for high-volume consumer electronics.

As reported by R. M. Emrick and J. L. Vokakis,¹² there is an excellent opportunity to use millimeter wave frequencies to propagate signals over short ranges supporting very high data rates from greater than 100 Mbps to 1–5 Gbps using available V-band spectrum at 60 GHz,

where there is 3 GHz of available unlicensed worldwide spectrum. Historically, the use of such high frequencies has been limited to specialized applications because of the high cost of implementation. Recent and ongoing advances in a number of technologies indicate that low-cost solutions utilizing these frequencies should be possible.¹³ Some early work indicates that Si technology transceivers can support the 60 GHz frequency band of operation and are economical enough to manufacture in large quantities for a large consumer market.^{14,15} Next-generation network interface cards (NIC) already support 1 Gbps transfer rates and USB2.0 and IEEE 1394 standards are in place for high-speed wireless capability, making the need for 60 GHz high-speed wireless hubs (or equivalent) more a question of when they will be needed rather than if they will be needed.

It is interesting to note that there is an overlap between the unlicensed bands worldwide in the V-band from 59–62 GHz, which will allow for 3 GHz of available bandwidth that can be readily used anywhere on the planet (see Figure 2-6). There is also a standard being developed in the form of IEEE 802.15.3c. Other spectrum purchases by supporting and competing 60 GHz technology companies are sure to occur outside this unlicensed band and will likely supplement the necessary bandwidth for a successful millimeter-wave high-speed broadband protocol.

Solutions utilizing this millimeter-wave spectrum for very high-speed connectivity may be “stand alone,” or they may be integrated with other wireless technologies where at least some of the system can be reused. Measured performance of an early proof-of-concept system is shown in Figure 2-7. In consumer electronics, typically acceptable bit error rates (BER) are around 10^{-6} . Using this BER, Figure 2-7 shows demonstrated data rates of about 3.5 Gbps at ranges of 5 m and 1 Gbps up to a range of 50 m. This performance is for point-to-point line-of-sight with a single antenna at the transmitter and receiver. In order to achieve acceptable performance for non-line-of-sight use, MIMO and antenna gain are most likely needed.¹²

As previously stated, the deployment of high-speed solutions using millimeter-wave frequencies can be stand alone or exist as an alternative mode in a multimode system that includes 802.11. The European Union funded a program called Broadway, which looked at a hybrid dual-frequency broadband wireless LAN access system at 5 GHz and 60 GHz to be implemented for urban hotspot coverage.¹⁶ This system is based on integrating HIPERLAN/2 OFDM technology at 5 GHz and an ad-hoc extension of it at 60 GHz called HIPERSHOT and would be an extension of the existing 5 GHz WLAN systems using new modes in the 59–65 GHz worldwide unlicensed frequency bands. Broadway could alternatively be used in combination with 802.11 solutions that exist and utilize the 2.4 GHz band.

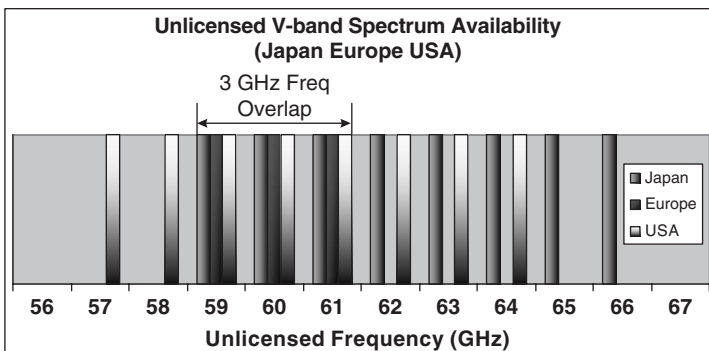


FIGURE 2-6 Available unlicensed global spectrum in the V-band

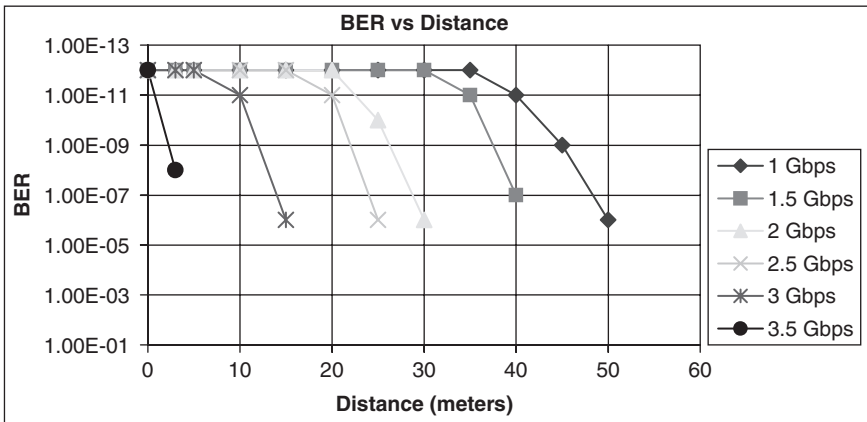


FIGURE 2-7 Measured data rate vs. distance for an early proof of concept system,¹³ which shows the potential for very high data rates at distances traditionally within the WPAN ranges for point-to-point line-of-sight (after R. Emrick et al¹³ © IEEE 2005)

2.4 EXAMPLES OF FUNDAMENTAL ANTENNA TYPES USED IN MILITARY AND COMMERCIAL APPLICATIONS

The selection of an antenna for a specific application is typically driven by a number of factors including antenna size, which is associated with the particular frequency being utilized and the platform it will fit, the direction and distance over which the antenna needs to transmit its signal, and the antenna's operational environment, which can introduce many variables that are best handled by certain antenna designs.

There are a number of fundamental antenna types that are typical options when determining the best antenna for a given application and frequency band. There are, of course, variations of these antenna types that may, at times, include hybrid approaches providing for optimized solutions. Many antennas, such as monopoles and dipoles, are resonant antennas that generally have dimensions close to half a wavelength. These antennas are often simple structures with relatively narrow bandwidth. Broadband antennas, such as log-periodic antennas, typically have the largest dimension that is a half-wavelength at the lowest frequency of operation. Antennas like this become resonant across the full frequency of operation for greatly improved bandwidth. Aperture antennas tend to be used for higher gain, are typically several wavelengths in one or more dimensions, and can be implemented using a feed and reflector or an array of wire or printed elements.

Dipole and monopole wire antennas are the most basic types of antenna because of their simple structure and omnidirectional radiation capability. The impedance characteristics associated with these antennas have been investigated thoroughly, giving rise to general formulas for straightforward design. A monopole is an antenna that is mounted ideally on an infinite ground plane and can be characterized by a dipole antenna that is twice its size radiating in free space. A simple half-wave dipole is a center-fed driven antenna whose two conductors are opposite each other and perpendicular to the center feed. There are many different variations of these antennas including the cylindrical dipole, which has a length to diameter ratio of more than 15:1, biconical dipoles with conductors that flare out into cones, folded dipole

antennas that are formed by two cylindrical dipoles and driven by a pair of transmission lines, and sleeved dipoles and monopoles, which are doubly fed antennas with the current at a maximum at the center of the dipole or base of the monopole. Their omnidirectional radiation patterns are well suited for terrestrial LOS communications over vast distances.

Reflector antennas have high gain and are very directive making them well suited for earth-based stations communicating with satellites in orbit. Reflector designs can be planar, have a parabolic arc, or incorporate a design that uses neither of the aforementioned designs. Feed structures can be placed in front of the reflector or be fed from the back, depending on the reflector technology being implemented. Reflectors have superior gain and directivity but require large apertures that can be quite cumbersome to position, requiring mechanical control and closed loop tracking, such as conical scanning when the ground station is mobile or there is satellite movement in a low earth orbit (LEO). The components needed with this technology take up a significant amount of space, creating the need for alternate solutions.

Phased array antennas utilize multiple radiator elements uniformly spaced in linear or two-dimensional configurations, often half a wavelength apart, and their amplitude and phases are controlled to form a radiated beam of any desired shape for propagation in free space. Phased arrays are normally planar, making them low profile and they produce high-gain beams that are electronically steered, making them a viable replacement for large reflector antennas. Phase shifters are most often the technology of choice to manipulate the phase between radiators. One of the biggest challenges in implementing a phased array is cost; each element of the array typically requires extensive and costly electronics. Historically, phased arrays have been used for less cost-sensitive applications, but future implementation using digital beam forming, retrodirection, and/or other alternative methods, may allow for more widespread use in more cost-sensitive applications.

Printed microstrip-type antennas are used in many different commercial applications, including cell phones, WLAN networking cards, personal digital assistants, and GPS receivers. These applications typically operate at ranges that are within 1–10,000 m of the transmission and reception points. These antennas have inherently low gain and, in most cases, require amplification to increase range, but the relative ease of fabrication and low profile make them attractive candidates for small commercial and military handheld communication devices that have limited space for antenna apertures.

In the end, the selection of the ideal antenna for a given frequency band and application comes down to balancing the fundamental parameters that typically include size, cost, weight, and reliability. Achieving the right balance among these parameters is often far from trivial.

REFERENCES

1. L. Yujiri, M. Shoucri, and P. Moffa, "Passive Millimeter-Wave Imaging," *IEEE Microwave Magazine* (September 2003): 39–50.
2. P. Ciaia, R. Staraj, G. Kossias, and C. Luxey, "Design Internal Quad-Band Antenna for Mobile Phones," *IEEE Microwave Wireless Components Letters*, vol. 14, no. 4 (April 2004): 148–150.
3. K. L. Wong and S. C. Pan, "Compact Triangular Microstrip Antenna," *Electron. Lett.*, 33 (March 13, 1997): 433–434.
4. R. Waterhouse, "Small Microstrip Patch Antenna," *Electron. Lett.*, 31 (April 13, 1995): 604–605.
5. S. Dey and R. Mittra, "Compact Microstrip Patch Antenna," *Microwave Opt. Technol. Lett.*, 13 (September 1996): 12–14.
6. K. L. Wong, *Compact and Broadband Microstrip Antennas*, (New York: Wiley, 2002).

7. Y. F. Lin and K. L. Wong, "Compact Broadband Triangular Microstrip Antennas with an inset Microstrip-line Feed," *Microswave Opt. Technol. Lett.*, 17 (February 20, 1998): 169–170.
8. J. H. Lu, C. L. Tang, and K. L. Wong, "Slot-Coupled Small Triangular Microstrip Antenna," *Microwave Opt. Technol. Lett.*, 16 (December 20, 1997): 371–374.
9. J. H. Lu and K. L. Wong, "Slot-loaded, Meandered Rectangular Microstrip Antenna with Compact Dual-frequency Operation," *Electron. Lett.*, 34 (May 28, 1998): 1048–1050.
10. C. K. Wu, K. L. Wong, and W. S. Chen, "Slot-coupled Meandered Microstrip Antenna for Compact Dual-frequency Operation," *Electron. Lett.*, 34 (May 28, 1998): 1047–1048.
11. M. Weiser, "Hot Topics: Ubiquitous Computing" *IEEE Computer*, (October 1993).
12. R. M. Emrick and J. L. Volakis, "Antenna Requirements Short Range High speed Wireless Systems Operating at Millimeter-wave Frequencies," 2006 IEEE IMS International Microwave Symposium, June 2006.
13. R. Emrick, S. Franson, J. Holmes, B. Bosco, and S. Rockwell, "Technology for Emerging Commercial Applications at Millimeter-wave Frequencies," IEEE/ACES International Conference on Wireless Communications and Applied Computational Electromagnetics, April 2005.
14. S. Reynolds, B. Floyd, U. Pfeiffer, and T. Zwick, "60 GHz Direct-conversion Transceiver Circuits in SiGe Bipolar Technology," *IEEE ISSCC Dig. Tech. Papers* (February 2004): 442–538.
15. C. Doan, S. EMami, D. Sobel, A. Niknejad, and R. Brodersen, "60GHz CMOS Radio for Gb/s Wireless LAN," *IEEE RFIC Sym. Dig Papers* (June 2004): 225–228.
16. M. de Courville, S. Zeisberg, Markus Muck, and J. Schönthier, "BROADWAY—the Way to Broadband Access at 60 GHz," IST Mobile & Wireless Telecommunications Summit, June 2002.

Chapter 3

Arrays of Discrete Elements

John N. Sahalos

Aristotle University of Thessaloniki

CONTENTS

3.1 INTRODUCTION.....	3-2
3.2 ANTENNA ARRAY FACTOR AND ANTENNA INDICES	3-2
3.3 LINEAR ARRAYS.....	3-4
3.4 PLANAR ARRAYS.....	3-13
3.5 THREE DIMENSIONAL AND CONFORMAL ARRAYS.....	3-16
3.6 ARRAY SYNTHESIS TECHNIQUES	3-16
3.7 SMART ANTENNAS.....	3-22
3.8 ELEMENT PATTERN AND MUTUAL COUPLING.....	3-23

3.1 INTRODUCTION

Telecommunications services in which antennas occupy a prominent position have grown tremendously over the years. Antennas are ubiquitous devices in radio, TV, satellite, and mobile communications, and the economic and social impact of these services motivate engineers to improve antenna feasibility and performance. A single element antenna, due to its limited performance, is often not enough to meet technical needs for high gain, narrow and/or steerable beams, pattern nulls, and low sidelobes. An array of discrete elements can, however, cover most of these constraints. The antenna array is characterized by element type, geometry, and the required excitations. Mutual coupling between the elements is important in the array design. The elements are chosen, if possible, to be identical and parallel. Uniformly spaced linear arrays are the most practical ones.

In this chapter the properties of the various antenna arrays and many synthesis methods will be discussed.

3.2 ANTENNA ARRAY FACTOR AND ANTENNA INDICES

The approximation of the far field coming from an antenna¹ separates it into a function of the position r of the observation point and into a function-vector of its angles, θ and ϕ , of the spherical coordinate system. The far electric field of a typical antenna array element (see Figure 3-1) can be expressed as^{2,3}

$$E_n(r) \equiv -j\omega\mu \frac{e^{-j\beta r}}{4\pi r} f_n(\theta, \phi) \quad (3-1)$$

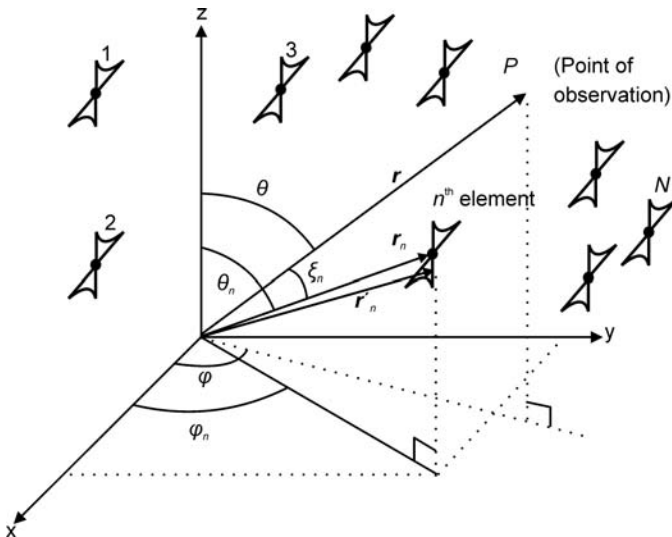


FIGURE 3-1 Geometry of a typical antenna array

The angular-dependent vector $f_n(\theta, \phi)$ presents the directional characteristics of the n th element.

Because the antenna array shown in Figure 3-1 contains N elements, the total electric field is

$$E(r) = \sum_{n=1}^N E_n(r) \quad (3-2)$$

As mentioned previously, the elements of the array are identical and, if possible, identically oriented. In this case, the current distribution of the elements is approximately the same. So, $f_n(\theta, \phi)$ is

$$f_n(\theta, \phi) = I_n f(\theta, \phi) \quad (3-3)$$

$f(\theta, \phi)$ characterizes the pattern of the element, and I_n is the excitation of the n th element of the array. By taking into account Eq. 3-1 to Eq. 3-3, the total electric field is

$$E(r) = -j\omega\mu \frac{e^{-j\beta r}}{4\pi r} f(\theta, \phi) \sum_{n=1}^N I_n e^{j\beta r_n \cos \xi_n} \quad (3-4)$$

which is $\cos \xi_n = \sin \theta \sin \theta_n \cos(\phi - \phi_n) + \cos \theta \cos \theta_n$, where (r_n, θ_n, ϕ_n) are the coordinates of a convenient reference point of the n th element. The summation of Eq. 3-4 is the *array factor* $AF(\theta, \phi)$:

$$AF(\theta, \phi) = \sum_{n=1}^N I_n e^{j\beta r_n \cos \xi_n} \quad (3-5)$$

$AF(\theta, \phi)$ actually presents the pattern of N isotropic point sources from the original array. Equation 3-5 then becomes

$$E(r) = -j\omega\mu \frac{e^{-j\beta r}}{4\pi r} f(\theta, \phi) AF(\theta, \phi) \quad (3-6)$$

Equation 3-6 states that, “An array of similar identically oriented elements has a pattern that can be expressed as the product of the element pattern and the array factor.” Since the element pattern is known, the main effort in antenna array design is the synthesis of the array factor.

One of the main characteristics of an antenna is its radiation pattern (see Figure 3-2). The pattern presents graphically the radiation properties that can be measured by moving a probe around the antenna when testing at a constant distance in the far field. If the probe moves in a plane, the pattern is known as the *plane pattern*. The plane pattern that contains the electric field vector is called the *E-plane pattern*. Correspondingly, the plane pattern that contains the magnetic field vector is the *H-plane pattern*.

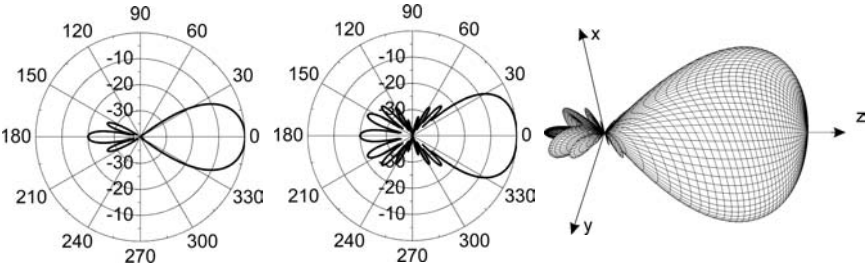


FIGURE 3-2 Plane and 3D patterns of an antenna

The radiation characteristics of an antenna array are measured in terms of the indices given in Chapter 1. These indices are the directivity $D(\theta_0, \phi_0)$, the gain $G(\theta_0, \phi_0)$, the signal-to-noise ratio SNR , the radiation efficiency η , the quality factor Q , the half power (or 3 dB power) beamwidth $HPBW$, the first null beamwidth BW_{null} , the sidelobe level SLL , and the bandwidth of the antenna array.

3.3 LINEAR ARRAYS

A practical array that consists of a number of elements set up along a straight line is called a *linear array* (see Figure 3-3). The pattern of a linear array can be expressed in well-known polynomial forms.

The array factor $AF(\theta, \phi)$ of the linear array shown in Figure 3-3 depends only on the angle θ and is written as

$$AF(\theta) = \sum_{n=0}^{N-1} I_n e^{j\beta d_n \cos \theta} \quad (3-7)$$

If the elements are equally spaced in terms of distance d , then Eq. 3-7 yields to

$$AF(\theta) = \sum_{n=0}^{N-1} I_n e^{j\beta n d \cos \theta} = \sum_{n=0}^{N-1} I_n z^n \quad (3-8)$$

$$z = e^{j\beta d \cos \theta} \quad (3-9)$$

For $0 \leq \theta \leq \pi$, $AF(\theta)$ is a polynomial of z , which moves on a unit circle.³ The phase is bounded between $-\beta d$ and $+\beta d$, and the bounded region is called the *visible region*.

Uniform Linear Arrays

Uniform linear arrays (ULA) consist of equally spaced elements with equal magnitude and progressive phases. In Eq. 3-7, $I_n = (e^{j\alpha})^n$. Equation 3-7 is then transformed to¹

$$AF(\psi) = e^{j(N-1)\psi/2} \frac{\sin(N\psi/2)}{N \sin(\psi/2)} \quad (3-10)$$

$$\psi = \beta d \cos \theta - \alpha \quad (3-11)$$

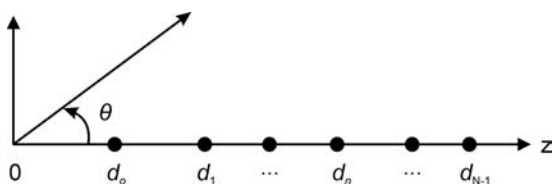


FIGURE 3-3 A linear array of N elements

The basic characteristics of the ULA array factor can be summarized as follows:

1. Maximum values occur at $\psi = \pm 2k\pi$, where $k = 0, 1, 2, \dots$
2. Nulls of the array are located at $\psi = \pm 2k\pi/N$, where $k = 1, 2, 3, \dots$ and $k \neq N, 2N, 3N, \dots$
3. The 3 dB points of the array factor occur when

$$AF(\psi) = \pm \frac{\sqrt{2}}{2} \Rightarrow |AF(\psi)|^2 = \frac{1}{2} \Rightarrow 10 \log |AF(\psi)|^2 = -3 \text{ dB}$$

which gives $N\psi/2 = \pm 1.391 \text{ rad}$.

4. Secondary maxima (minor lobes) occur when

$$\psi = (2k+1)\pi/N, \quad \text{where } k = 1, 2, 3, \dots$$

The first minor lobe is located at $\psi = \pm 3\pi/N$ with

$$|AF(\psi)| = \frac{1}{N \sin \frac{3\pi}{2N}} = SLL$$

For a phase shift $\alpha = \beta d \cos \theta_0$, the array beam can be steered in the direction θ_0 . For an end-fire array, $\alpha = \mp \beta d$. Apart from the ordinary end-fire array, there is also the Hansen-Woodyard (HW) array, which is more directive.² With HW arrays, the progressive phase shift α is

$$\alpha = \mp \left(\beta d + \frac{2.94}{N} \right) = \mp \left(\beta d + \frac{\pi}{N} \right) \quad (3-12)$$

For HW arrays,

$$|\psi| = \pi \begin{cases} \text{for } \theta = \pi \text{ if maximum occurs at } \theta = 0 \\ \text{for } \theta = 0 \text{ if maximum occurs at } \theta = \pi \end{cases} \quad (3-13)$$

$$d = \frac{\lambda}{4} \left(1 - \frac{2.94}{\pi N} \right) = \frac{\lambda}{4} \left(1 - \frac{1}{N} \right) \quad (3-14)$$

Table 3-1 presents the sidelobe locations and levels for broadside ULAs with different numbers of elements. The directivity of broadside ULAs is presented in Table 3-2.

A linear array with maximum at $\theta = \theta_0$ and half-power beamwidth θ_H must have

$$\theta_0 > \cos^{-1} \left(\cos^2 \frac{\theta_H}{2} \right) \quad (3-15)$$

$$N \frac{d}{\lambda} = 0.4428 \left[\frac{2(1 + \cos \theta_H)}{|\sin^4 \theta_0 - (\cos \theta_H - \cos^2 \theta_0)|^2} \right]^{1/2} \quad (3-16)$$

In end-fire arrays where $\theta_0 = 0^\circ$, Eq. 3-16 is modified to

$$N \frac{d}{\lambda} = \frac{0.4428}{1 - \cos \frac{\theta_H}{2}} \quad (\text{Ordinary}) \quad (3-17)$$

$$N \frac{d}{\lambda} = \frac{0.1398}{1 - \cos \frac{\theta_H}{2}} \quad (\text{Hansen - Woodyard}) \quad (3-18)$$

TABLE 3-1 Sidelobe Positions and Levels for Broadside ULAs of Isotropic Elements

N	Sidelobe							
	First		Second		Third		Fourth	
	Position (ψ_1)	Level (dB)	Position (ψ_2)	Level (dB)	Position (ψ_3)	Level (dB)	Position (ψ_4)	Level (dB)
4	± 131.81	-11.30						
5	± 104.48	-12.04	± 180.00	-13.98				
6	± 86.66	-12.43	± 149.12	-15.25				
7	± 74.08	-12.65	± 127.42	-15.98	± 180.00	-16.90		
8	± 64.71	-12.80	± 111.28	-16.43	± 157.14	-17.89		
9	± 57.45	-12.90	± 98.79	-16.73	± 139.48	-18.54	± 180.00	-19.08
10	± 51.66	-12.97	± 88.84	-16.95	± 125.41	-18.99	± 161.82	-19.89
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
N large	$\pm 540/N$	-13.46	$\pm 900/N$	-17.90	$\pm 1260/N$	-20.82	$\pm 1620/N$	-23.01

Chebyshev Arrays

Uniformly spaced linear arrays with nonuniform excitation of the elements can use Chebyshev polynomials.³ These arrays produce the desired *SLL* and are called *Chebyshev arrays*.

A Chebyshev polynomial $T_m(x)$ of m th order and an independent variable x is an orthogonal polynomial. In region $-1 \leq x \leq 1$, it contains equal ripples with amplitudes between $+1$ and -1 . $T_m(x)$ outside the region $(-1, +1)$ rises exponentially. $T_m(x)$ is actually expressed as

$$T_m(x) = \begin{cases} \cos(m \cos^{-1} x) & |x| \leq 1 \\ \left(\frac{x}{|x|}\right)^m \cosh(m \cosh^{-1} |x|) & |x| > 1 \end{cases} \quad (3-19)$$

TABLE 3-2 Directivities for Broadside ULAs of Isotropic Elements

N	Spacing (λ)							
	0.125	0.250	0.375	0.500	0.625	0.750	0.875	1.000
4	1.28	2.16	3.11	4.00	4.84	5.58	5.29	4.00
5	1.45	2.70	3.83	5.00	6.12	6.97	7.68	5.00
6	1.67	3.17	4.63	6.00	7.30	8.54	9.52	6.00
7	1.93	3.64	5.34	7.00	8.61	10.10	11.22	7.00
8	2.20	4.16	6.10	8.00	9.84	11.55	12.82	8.00
9	2.48	4.68	6.86	9.00	11.07	12.99	14.42	9.00
10	2.74	5.17	7.59	10.00	12.41	14.43	16.10	10.00

For zero and first order as well as for the recursion relation, you have

$$\left. \begin{aligned} T_0(x) &= 1, T_1(x) = x \\ T_m(x) &= 2xT_{m-1}(x) - T_{m-2}(x) \end{aligned} \right\} \quad (3-20)$$

These equations are used to create the Chebyshev polynomials of any order.

Dolph-Chebyshev Arrays Dolph⁴ has found that maximum directivity for a given SLL can be obtained using Chebyshev polynomials. Their equal ripples describe the sidelobes, and the exponential increase beyond $|x| = 1$ gives the main lobe. The excitation distribution is symmetrical in the centre of the array (see Figure 3-4).

The independent variable of the Chebyshev polynomial is

$$x = x_0 \cos(\psi/2) \quad (3-21)$$

At $x = x_0$, the Chebyshev polynomial takes its maximum value R :

$$T_m(x_0) = R \quad \text{or} \quad x_0 = \cosh\left(\frac{1}{m} \cosh^{-1} R\right) \quad (3-22)$$

Nulls of $T_m(x)$ are located at

$$x_k = \pm \cos \frac{(2k-1)\pi}{2m}, \quad \psi_k = \pm 2 \cos\left(\frac{x_k}{x_0}\right) \quad \text{where } k = 1, 2, \dots, m \quad (3-23)$$

By using the expression $z_k = e^{j\psi_k}$, you can find I_k from the following polynomial expression⁵:

$$AF(\theta) = C \prod_{n=1}^m (z - z_n) = C \sum_{k=0}^m I_k z^k \quad (3-24)$$

The order of the Chebyshev polynomial is equal to the total number of array elements minus one.

An example of two patterns (broadside and intermediate) of a 14-element array with $SLL = -25$ dB is given in Figure 3-5. The polynomial is $T_{13}(x)$. For $R = 25$ dB, it is $R = 10^{25/20} = 17.7828$ and $x_0 = 1.038$.

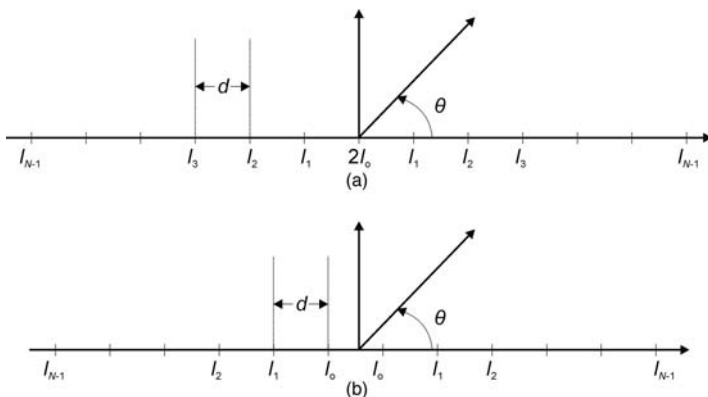


FIGURE 3-4 The geometry of an even and an odd element linear array

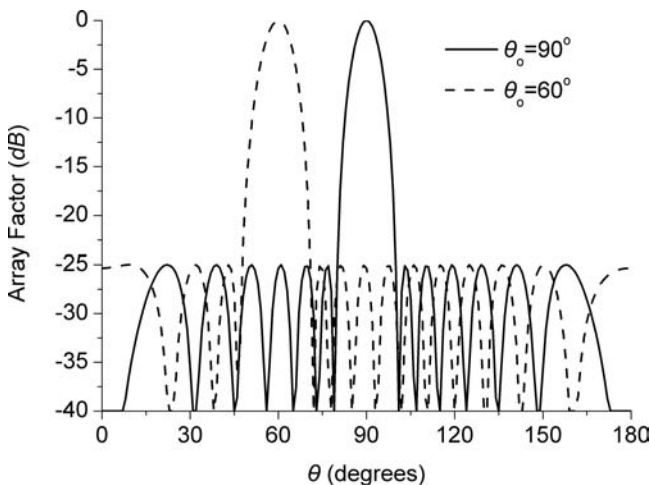


FIGURE 3-5 Pattern of a 14 element Dolph-Chebyshev array with $SLL = -25$ dB and $d/\lambda = 0.5$. (a) broadside array $\theta_0 = 0^\circ$, (b) intermediate array with maximum at $\theta_0 = 60^\circ$.

In Chebyshev arrays, there is a maximum permissible interelement distance d_{\max} to avoid grating lobes:

$$\frac{d_{\max}}{\lambda} = 1 - \frac{\cos^{-1}(1/x_0)}{\pi} \quad (3-25)$$

x_0 is the distance where the maximum of the Chebyshev polynomial occurs.

The ratio of the Dolph array's $HPBW$ over a uniform array ($HPBW_u$) of the same length gives the broadening factor f , which is found to be¹

$$f = \frac{(HPBW)}{(HPBW_u)} = 1 + 0.632 \left[\frac{2}{R} \cosh \sqrt{(\cosh^{-1} R)^2 - \pi^2} \right]^2 \quad (3-26)$$

f is valid in the range of $-60 \text{ dB} \leq SLL \leq -20 \text{ dB}$ and for scanning near broadside arrays.

Estimation of the directivity D with the help of the factor f is of the form:¹

$$D = \frac{2R^2}{1 + (R^2 - 1)f \frac{\lambda}{Nd}} \quad (3-27)$$

For the general case, the $HPBW$ is given approximately by

$$HPBW \cong 10.3^\circ \frac{\lambda}{Nd} \sqrt{s + 4.52 \csc \theta_0} \quad (3-28)$$

s is the absolute value of the SLL in dB, and θ_0 is the scan angle. Also, the directivity D is

$$D = \frac{2R^2}{1 + R^2 \frac{\lambda}{Nd} \sqrt{\frac{\ln(2R)}{\pi}} \cdot \sin\left(\beta \frac{d}{2}\right)} \quad (3-29)$$

General Chebyshev Arrays Dolph's method is suitable for $d \geq \lambda/2$, but it fails to give the optimum design for $d < \lambda/2$. To solve the problem Riblet⁶ devised a Chebyshev method by using odd numbers of elements. Moreover in J. N. Sahalos' *Orthogonal Methods for Array Synthesis, Theory and the ORAMA Computer Tool*,³ a method is presented that covers Riblet's method. In this method, the independent variable is expressed by

$$x = a \cos(\beta d \cos \theta + \alpha) + b \quad (3-30)$$

The coefficients a , b , and α are found from three different angles, θ_1 , θ_2 and θ_3 , of the pattern that corresponds to three values of x , the x_1 , x_2 , and x_3 . The visible region is in $\min(x_1, x_2, x_3) \leq x \leq \max(x_1, x_2, x_3)$.

$$\left. \begin{aligned} x_1 &= a \cos(\beta d \cos \theta_1 + \alpha) + b \\ x_2 &= a \cos(\beta d \cos \theta_2 + \alpha) + b \\ x_3 &= a \cos(\beta d \cos \theta_3 + \alpha) + b \end{aligned} \right\} \quad (3-31)$$

From Eq. 3-31, you then have

$$\tan \alpha = \frac{\sin y_{21} - \lambda \sin y_{31}}{\lambda \cos y_{31} - \cos y_{21}} \quad (3-32)$$

$$\left. \begin{aligned} y_{21} &= \beta \frac{d}{2} (\cos \theta_2 + \cos \theta_1) \\ y_{31} &= \beta \frac{d}{2} (\cos \theta_3 + \cos \theta_1) \\ \lambda &= \frac{(x_2 - x_1) \sin \left[\frac{\beta d}{2} (\cos \theta_3 - \cos \theta_1) \right]}{(x_3 - x_1) \sin \left[\frac{\beta d}{2} (\cos \theta_2 - \cos \theta_1) \right]} \end{aligned} \right\} \quad (3-33)$$

$$a = \frac{x_2 - x_1}{\cos(\beta d \cos \theta_2 + \alpha) - \cos(\beta d \cos \theta_1 + \alpha)} \quad (3-34)$$

$$b = x_1 - a \cos(\beta d \cos \theta_1 + \alpha) \quad (3-35)$$

I_k is found by using a polynomial expression as in Eq. 3-24.⁵ For $(2m - 1)$ elements, $T_m(x)$ is used. Nulls of $T_m(x)$ are located at

$$x_k = \pm \frac{(2k-1)\pi}{2m}, \quad \psi_k = \pm \cos^{-1} \left(\frac{x_k - b}{a} \right) \quad k = 1, 2, \dots \quad (3-36)$$

The pattern of an 11 element broadside array with $SLL = -20$ dB and $d/\lambda = 0.3$ is shown in Figure 3-6. $T_5(x)$ is used and the present method offers a more directive pattern than the Dolph array. More examples of the design of General Chebyshev patterns can be found in J. N. Sahalos.^{3,7,8}

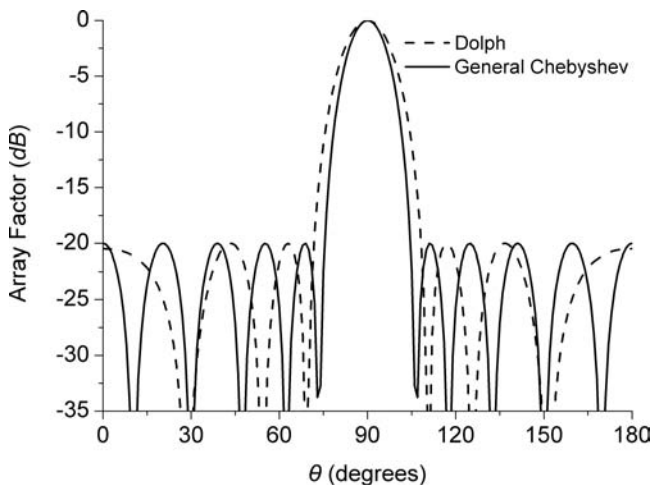


FIGURE 3-6 Pattern of an 11 element Dolph and General Chebyshev array with $SLL = -20$ dB and $d/\lambda = 0.3$

Taylor Distributions

Continuous distributions are useful when creating patterns with low sidelobes.¹ A discrete array coming from the appropriate sampling of a continuous distribution (line source) can give a similar pattern. Root matching is a suitable technique that offers satisfactory results. In root matching, the nulls of the patterns of the continuous distribution and the discrete array must be the same. If the patterns do not match with the desired accuracy, a perturbation technique⁹ in the discrete-element array is applied.

For a continuous line source, the array factor is called a *space factor* (SF). If L is the length, $I(z')$ and α are the amplitude distribution and phase progress along the line source, respectively, SF is expressed by

$$SF(\theta) = \int_{-L/2}^{L/2} I(z') e^{j(\beta \cos \theta - \alpha) z'} dz' = \int_{-L/2}^{L/2} I(z') e^{j\xi z'} dz' = \int_{-\infty}^{\infty} I(z') e^{j\xi z'} dz' \quad (3-37)$$

$I(z')$ is zero outside of $-L/2 \leq z' \leq L/2$. From Eq. 3-37, by using the Fourier transform, you obtain

$$I(z') = \frac{1}{2\pi} \int_{-\infty}^{\infty} SF(\theta) e^{-j\xi z'} d\xi \quad (3-38)$$

The desired $SF(\theta)$ needs a distribution that exists in $-\infty \leq z' \leq \infty$. However, if L is large enough, Eq. 3-38 can be used for the current distribution up to $(-L/2, L/2)$ within a certain error. Taylor gave a distribution, with an optimally low SLL for the first N sidelobes next to the main beam. Sidelobes, after the first N , gradually fall off in value. The SF of the Taylor

distribution comes from the Dolph-Chebyshev array if the elements of the array become infinite.¹⁰ It is

$$SF(u) = \frac{\cosh\left[\sqrt{(\pi A)^2 - u^2}\right]}{\cosh(\pi A)} \quad (3-39)$$

$$\left. \begin{aligned} u &= \pi \frac{L}{\lambda} (\cos \theta - \cos \theta_0) \\ \cosh(\pi A) &= R \end{aligned} \right\} \quad (3-40)$$

In practice, Taylor's distribution¹¹ is given by a space factor whose roots are the zeros of $SF(u)$. This distribution is called the *Chebyshev error distribution*. In addition, Taylor synthesized a pattern that has the first sidelobes at a certain level with the others decaying as the scan angle increases. Enough samples in the distribution may synthesize an array with its pattern having approximately similar characteristics.

Bayliss Distributions

Bayliss distributions¹² create patterns with their null along the boresight and the sidelobes at a given level. Bayliss' space factor is

$$SF(u) = u \cosh(\pi u) \frac{\prod_{n=1}^{\bar{n}-1} \left[1 - \left(\frac{u}{u_n} \right)^2 \right]}{\prod_{n=1}^{\bar{n}-1} \left[1 - \left(\frac{u}{n + \frac{1}{2}} \right)^2 \right]} \quad (3-41)$$

In this, $(\bar{n} - 1)$ defines the number of inner nulls, and u_n is given as a function of two quantities A and ξ_n .

$$u_n = \begin{cases} \left(\bar{n} + \frac{1}{2} \right) \left(\frac{\xi_n^2}{A^2 + \bar{n}^2} \right)^{1/2} & n = 1, 2, 3, 4 \\ \left(\bar{n} + \frac{1}{2} \right) \left(\frac{A^2 + n^2}{A^2 + \bar{n}^2} \right)^{1/2} & n = 5, 6, \dots, \bar{n} - 1 \end{cases} \quad (3-42)$$

Modified Taylor and Bayliss Patterns

Patterns based on the Taylor and Bayliss distributions can be modified to have their sidelobes at different levels. The space factors then take the more general form.⁵

(i) Taylor

$$SF(u) = C_0 \frac{\sin \pi u}{u} \frac{\prod_{n=\bar{n}_L-1}^{\bar{n}_R-1} \left(1 - \frac{u}{u_n} \right)}{\prod_{n=\bar{n}_L-1}^{\bar{n}_R-1} \left(1 - \frac{u}{n} \right)} \quad (3-43)$$

$$\left. \begin{aligned} u_n &= \bar{n}_R \frac{\sqrt{A_R^2 + \left(n - \frac{1}{2}\right)^2}}{\sqrt{A_R^2 + \left(\bar{n}_R - \frac{1}{2}\right)^2}} \\ u_n &= -\bar{n}_L \frac{\sqrt{A_L^2 + \left(n + \frac{1}{2}\right)^2}}{\sqrt{A_L^2 + \left(\bar{n}_L - \frac{1}{2}\right)^2}} \end{aligned} \right\} \quad (3-44)$$

in which \bar{n}_R and \bar{n}_L denote the transition roots of the right and left side, respectively, and A_R and A_L are the corresponding *SLL* parameters.

(ii) Bayliss

$$SF(u) = C_0 u \cos \pi u \frac{\prod_{n=-(\bar{n}_L-1)}^{\bar{n}_R-1} \left(1 - \frac{u}{u_n}\right)}{\prod_{n=-\bar{n}_L}^{\bar{n}_R-1} \left(1 - \frac{u}{n + \frac{1}{2}}\right)} \quad (3-45)$$

To synthesize the distribution, you start from an initial space factor $SF_0(\theta)$ with the *SLLs* on both sides being the average of the desired ones. In this case, the roots of Eq. 3-43 and Eq. 3-45, u_n^0 , are known, and you assume that

$$u_n = u_n^0 + \delta u_n \quad (3-46)$$

with δu_n being a small perturbation of u_n^0 . For $C = C_0 + \delta C$, the new space factor $SF(u)$ is then of the form

$$\frac{SF(u)}{SF_0(u)} - 1 = \frac{\delta C}{C_0} + \sum_{n=-(\bar{n}_L-1)}^{\bar{n}_R-1} \frac{\frac{u}{(u_n^0)^2}}{1 - \frac{u}{u_n^0}} \delta u_n \quad (3-47)$$

If the peak positions are u_m^p , then

$$\frac{SF(u_m^p)}{SF_0(u_m^p)} - 1 = \frac{\delta C}{C_0} + \sum_{n=-(\bar{n}_L-1)}^{\bar{n}_R-1} \frac{\frac{u_m^p}{(u_n^0)^2}}{1 - \frac{u_m^p}{u_n^0}} \delta u_n \quad (3-48)$$

With $\bar{n}_R + \bar{n}_L - 1$ being the number of lobes, this leads to a system of linear equations similar to the one in Eq. 3-48. The system is solved for $\delta C/C_0$ and for the $\bar{n}_R + \bar{n}_L - 2$ values δu_n ($\delta u_0 = 0$). The new values of u_n produce a pattern comparable to the desired one. If the desired and new patterns differ more than the minimum predefined amount, the process is repeated until the desired tolerance is achieved.

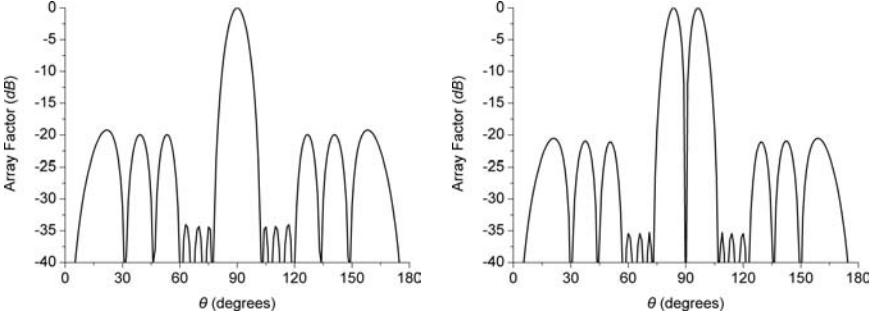


FIGURE 3-7 Patterns of (a) $N = 22$ elements based on the modified Taylor distribution (left), and (b) of $N = 23$ elements based on the modified Bayliss distribution; $d/\lambda = 0.5$, $\bar{n} = 6$, $SLL = -20$ dB with the three innermost lobes at -40 dB

For the Bayliss distribution you have

$$\frac{SF(u_m^p)}{SF_0(u)} - 1 = \frac{\delta C}{C_0} - \frac{\delta u_0}{u_m^p} + \sum_{n=-(\bar{n}_L-1)}^{\bar{n}_R-1} \frac{\frac{u_m^p}{(u_n^0)^2}}{1 - \frac{u_m^p}{u_n^0}} \delta u_n \quad (3-49)$$

This equation produces a system of $\bar{n}_R + \bar{n}_L$ unknowns, and the perturbation process is created in the same way as before. Sampling the distributions of the space factors, you obtain the corresponding arrays of discrete elements.

Figure 3-7 shows the patterns of two arrays taken by sampling two modified distributions with $\bar{n} = 6$, $SLL = -20$ dB, with the three innermost pairs of lobes at -40 dB.

3.4 PLANAR ARRAYS

Planar arrays are created from individual radiators positioned on a plane. They provide more variables and can give stricter constraints than linear ones. Also, they allow for three-dimensional mainbeam scanning. Planar arrays with their elements along a rectangular grid create rectangular arrays. In essence, two perpendicular linear arrays create a rectangular one. The array factor of an $M \times N$ rectangular array is

$$F(\theta, \phi) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_{mn} z_x^m z_y^n \quad (3-50)$$

In this, I_{mn} refers to the element excitations corresponding to the m th row and n th column. Also,

$$z_x = e^{j\beta d_x \sin\theta \cos\phi} \quad \text{and} \quad z_y = e^{j\beta d_y \sin\theta \sin\phi} \quad (3-51)$$

For an array with uniform excitation, the array factor is

$$|AF(\theta, \phi)| = \left| \frac{\sin(M\Psi_x/2)}{M \sin(\Psi_x/2)} \right| \left| \frac{\sin(N\Psi_y/2)}{N \sin(\Psi_y/2)} \right| \quad (3-52)$$

$$\left. \begin{aligned} \Psi_x &= \beta d_x \sin \theta \cos \phi + \alpha_x \\ \Psi_y &= \beta d_y \sin \theta \sin \phi + \alpha_y \end{aligned} \right\} \quad (3-53)$$

If d_x and/or d_y is $> \lambda$, grating lobes are created. They are produced when the in-phase addition of radiated fields happens in more than one direction. The main lobe and the grating lobes for planar arrays are located at

$$\Psi_x = \pm 2m\pi, \Psi_y = \pm 2n\pi, \quad m \text{ and } n = 0, 1, 2, \dots \quad (3-54)$$

α_x and α_y are independent to each other. From Eq. 3-53 and Eq. 3-54, the grating lobe direction (θ_{mn}, ϕ_{mn}) is

$$\tan \phi_{mn} = \left[\frac{\sin \theta_0 \sin \phi_0 \pm n\lambda/d_y}{\sin \theta_0 \cos \phi_0 \pm m\lambda/d_x} \right] \quad (3-55)$$

$$\sin \theta_{mn} = \left[\frac{\sin \theta_0 \sin \phi_0 \pm n\lambda/d_y}{\sin \phi_{mn}} \right] \quad (3-56)$$

The main lobe occurs for $m = n = 0$ where $\theta_{00} = \theta_0$ and $\phi_{00} = \phi_0$. A 3D pattern for 8×6 elements with $\alpha_x = \alpha_y = 0$ and $d_x = d_y = \lambda$ is given in Figure 3-8a. Due to large spacing, except for the maximum at $\theta_0 = 0$ and $\theta_0 = \pi$, grating lobes are created at $\theta = \pi/2$ and $\phi = 0, \pi/2, \pi, 3\pi/2$.

Rectangular arrays can be designed by also using nonuniform linear arrays. In Figure 3-8b the pattern of two Dolph-Chebyshev arrays with $SLL = -20$ dB for 8×6 elements with $d_x = d_y = \lambda/2$ is presented.

Circular Arrays

A planar array with the elements positioned in a circular ring creates a circular array. Figure 3-9 presents N isotropic elements spaced on a circular ring with radius R .

The array factor is

$$AF(\theta, \phi) = \sum_{n=1}^N I_n e^{j\beta R [\sin \theta \cos(\phi - \phi_n) - \sin \theta_0 \cos(\phi_0 - \phi_n)]} \quad (3-57)$$

I_n is the amplitude of the excitation and (θ_0, ϕ_0) is the direction where maximum occurs.

A short dipole positioned at the bisector of a corner reflector with angle $\omega_N = \pi/N$ produces $2N - 1$ images that create a circular array. The radiation pattern of this array can be derived from the contribution of the element and its images. The real pattern is in front of the corner.

M circular arrays in concentric rings have an array factor of

$$AF(\theta, \phi) = \sum_{m=1}^M \sum_{n=1}^N I_{mn} e^{j[\beta R_m \sin \theta \cos(\phi - \phi_{mn}) + \alpha_{mn}]} \quad (3-58)$$

$I_{mn} e^{j\alpha_{mn}}$ presents the excitation of the n th element of the m th ring.

A usual concentric ring can come from a corner reflector with a linear array positioned in front of it.

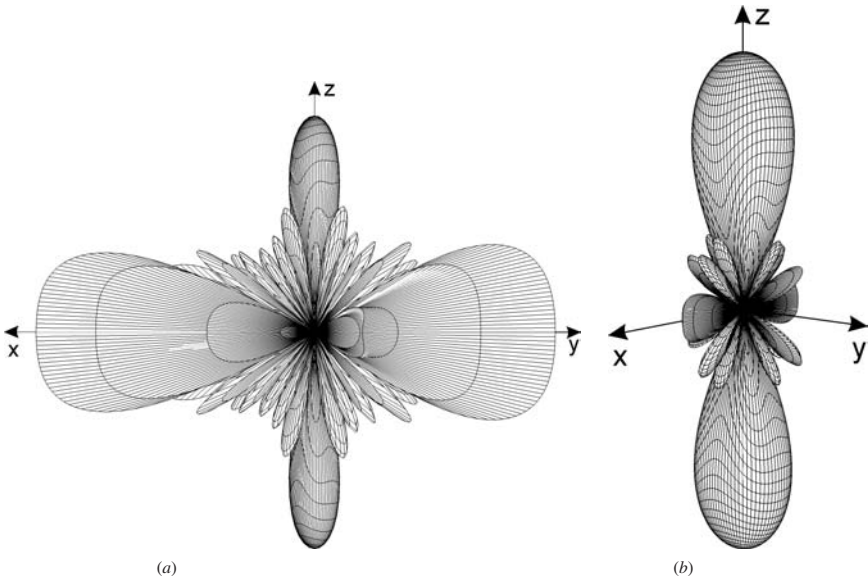


FIGURE 3-8 Three-dimensional patterns for an 8×6 rectangular array with $\alpha_x = \alpha_y = 0$: (a) Uniform arrays with $d_x = d_y = \lambda$; (b) Dolph-Chebyshev arrays with $SLL = -20$ dB and $d_x = d_y = \lambda/2$

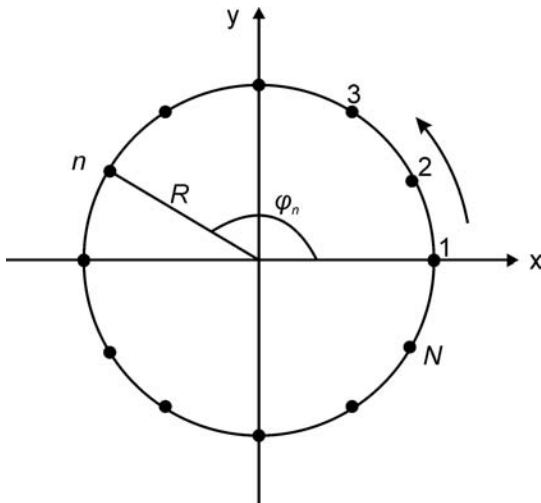


FIGURE 3-9 Geometry of a circular array with radius R and N isotropic sources

3.5 THREE DIMENSIONAL AND CONFORMAL ARRAYS

A 3D array (see Figure 3-1) with N elements has an array factor of the form

$$AF(\theta, \phi) = \sum_{n=1}^N I_n e^{j[\alpha_n + \beta r_n [\sin \theta \sin \theta_n \cos(\phi - \phi_n) + \cos \theta \cos \theta_n]]} \quad (3-59)$$

$I_n e^{j\alpha_n}$ is the complex excitation, and (r_n, θ_n, ϕ_n) are the spherical coordinates of the n th element. A *cylindrical array* is a special kind of 3D array. Circular arrays of the same radius with their centers on the same axis create a cylindrical array. A linear array in front of a corner reflector of angle $\omega_N = \pi/N$ with its axis parallel to the edge positioned at its bisector is a virtual cylindrical array. Such an array is an extremely useful 3D array with applications in radio and mobile communications.

Conformity requirements of arrays to a shaped surface create *conformal arrays*. Conformal arrays are used in mobile platforms and in stationary shaped surfaces for specified angles of coverage.

The pattern of a conformal array is given by Eq. 3-1 and Eq. 3-2 because the element pattern and the array factor are inseparable. Surface illumination, polarization, and the pattern of each element have to be taken into account separately. Cylindrical, conical, and parabolic conformal arrays are special and suitable for missiles and aircrafts. An example of a cylindrical array of microstrip patches and its pattern that has $SLL < -30$ dB and maximum at $\theta_0 = \pi/2$ and $\phi_0 = 0$ are presented in Figure 3-10.

3.6 ARRAY SYNTHESIS TECHNIQUES

Synthesis of an array aims at the approximation of a desired radiation pattern with a given accuracy. Several analytical techniques of narrow-beam and low SLL patterns have been presented previously. Moreover, the mean square error, the orthogonal method, the matrix

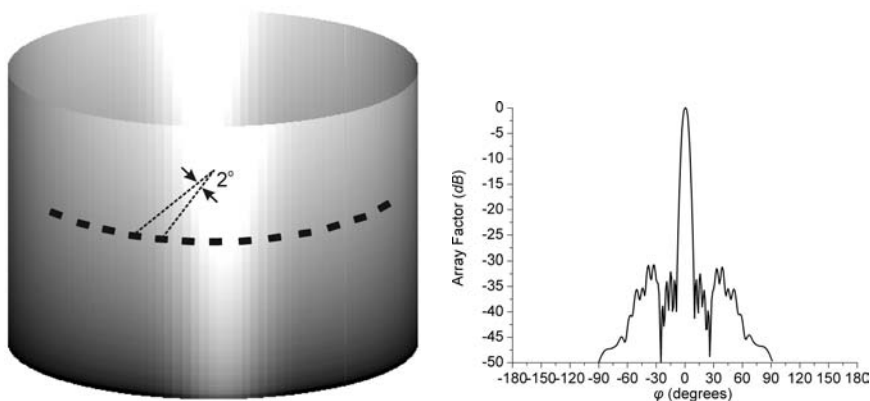


FIGURE 3-10 (a) A cylindrical conformal array with rectangular patches in interelement distance 1.16λ ; (b) The pattern for $N = 24$ elements and $SLL < -30$ dB

method, the simplex, the gradient method, the simulated annealing, and the Genetic algorithms are used for the pattern synthesis.

Orthogonal Methods

The orthogonal methods transform the basis of the array factor to an orthogonal basis. The new basis can easily handle the synthesis problem by deriving the excitation of the array. If the geometry of an array has to be found, the orthogonal perturbation method is applied. In this case, the position of the array elements, in an orthogonal manner, is perturbed, and the final geometry is derived by iteration procedure.

The Woodward-Lawson (WL) method is a simple orthogonal technique where an orthogonal set of beams is used.¹ Also, Fourier transform is a method based on the orthogonal components of the array factor.

Woodward-Lawson Method

The Woodward-Lawson (WL) method for discrete arrays can be implemented by superimposing groups of beams of uniform linear arrays. A uniform linear array with N elements of equal distance d/λ creates a beam pattern of the form

$$f_m(\theta) = b_m \frac{\sin(N\psi_m/2)}{N \sin(\psi_m/2)} \quad (3-60)$$

$$\psi_m = \beta d (\cos \theta - \cos \theta_m) \quad (3-61)$$

The superposition of $2M + 1$ terms of the form in Eq. 3-60 gives an array factor that is

$$AF(\theta) = \sum_{m=-M}^M b_m \frac{\sin(N\psi_m/2)}{N \sin(\psi_m/2)} \quad (3-62)$$

Sampling $AF(\theta)$ at the angles

$$\theta_m = \cos^{-1} \left(m \frac{\lambda}{Nd} \right) \quad (3-63)$$

we have

$$b_m = AF(\theta_m) \quad (3-64)$$

The excitation coefficients of a linear array with an array factor of the form of Eq. 3-62 are

$$I_n = \frac{1}{N} \sum_{m=-M}^M AF(\theta_m) e^{-j\beta d_n \cos \theta_m} \quad (3-65)$$

If, instead of a linear array, you have a line source, you can superimpose groups of beams of uniform line sources. In this case, you create a space factor with similar characteristics to the array factor.

A discrete element array can be designed from the line source by sampling the previous distribution. An example for a cosecant-squared power pattern by sampling the appropriate line source will be given next. Figure 3-11 presents the pattern by a line source with $L = 10\lambda$ and the same pattern of a discrete linear array with 20 elements. The excitation of the array is coming from the uniform sampling of the line source.

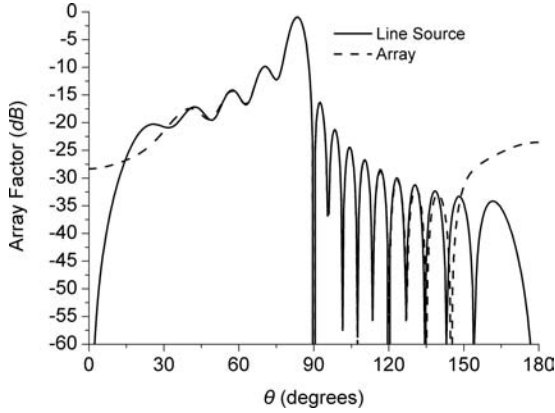


FIGURE 3-11 Cosecant-squared power pattern of a line source with $L = 10\lambda$ and of a discrete linear array with 20 elements. The array is coming from the uniform sampling of the line source.

Fourier and Orthogonal Method

A nonuniformly spaced linear array (see Figure 3-3) gives an array factor of the form

$$AF_d(u) = \sum_{n=1}^N I_n e^{jx_n u} \quad (3-66)$$

$$u = \pi \cos \theta, \quad x_n = \frac{d_n}{\lambda/2} \quad (3-67)$$

In general, the expansion functions $e^{jx_n u}$ are not orthogonal because the inner product $K_{in} \neq 0$:

$$K_{in} = \int_{-\pi}^{\pi} e^{j(x_i - x_n)u} du = \frac{\sin(x_i - x_n)\pi}{(x_i - x_n)\pi} \neq 0 \quad (3-68)$$

$AF_d(u)$ can be expressed in an orthogonal basis $\{\Psi_n(u)\}$ by the Gram-Schmidt procedure,³ which is

$$\Psi_n(u) = \sum_{i=1}^n c_i^{(n)} e^{jx_i u} \quad (3-69)$$

$$AF_d(u) = \sum_{n=1}^N B_n \Psi_n(u) \quad (3-70)$$

From Eq. 3-70, you then have

$$B_n = \int_{-\pi}^{\pi} AF_d(u) \cdot \Psi_n^*(u) du \quad (3-71)$$

$\Psi_n^*(u)$ is the conjugate of $\Psi_n(u)$. Combining Eq. 3-66 to Eq. 3-70, you find that

$$I_n = \sum_{i=n}^N c_n^{(i)} B_i \quad (3-72)$$

Equation 3-72 shows that the excitation of the elements can be derived by using the orthogonal method. If the elements of the array are uniformly spaced with $K_{in} = 0$, then the previous method is reduced to the Fourier. Several applications that explain the orthogonal procedure can be found in the literature.³

An example of a 13-element array with 0.87λ spacing and a constant beam between 87° and 93° is given here. Figure 3-12 presents this pattern. The orthogonal method can be used for 2D and 3D arrays with a similar formulation. Also, with modification, it can take into account the coupling between the elements.

Orthogonal Perturbation Method

The orthogonal perturbation method combines an iterative technique with the orthogonal method. A nonuniformly spaced linear array has an array factor of the form

$$AF_0(\theta) = \sum_{i=1}^N I_i e^{j\beta d_i \cos \theta} = \sum_{i=1}^N I_i \Phi_i(\theta) \quad (3-73)$$

A perturbation of the position d_i of each element such that $\beta(\delta d_i) \ll 1$ transforms the array factor to

$$AF_1(\theta) \cong \sum_{i=1}^N [1 + j\beta(\delta d_i) \cos \theta] I_i e^{j\beta d_i \cos \theta} \quad (3-74)$$

From Eq. 3-73 and Eq. 3-74 you can derive

$$F(\theta) = \frac{AF_1(\theta) - AF_0(\theta)}{\cos \theta} = \sum_{i=1}^N A_i \Phi_i(\theta) \quad (3-75)$$

$$A_i = j\beta(\delta d_i) I_i \quad (3-76)$$

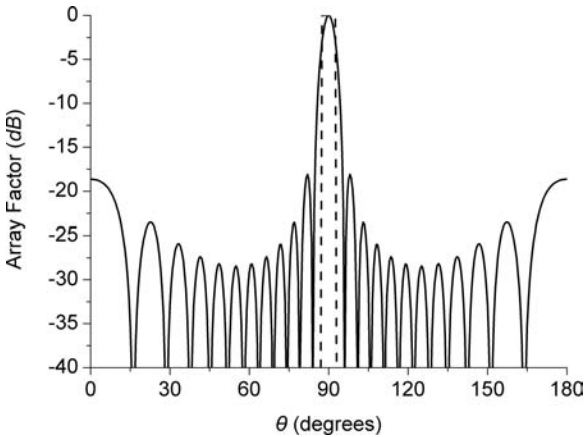


FIGURE 3-12 Beam from 87° to 93° by the orthogonal method of an array with $N=13$ and $d/\lambda = 0.87\lambda$

$F(\theta)$ is similar to $AF_0(\theta)$, and the orthogonal method to establish an orthogonal basis $\{\Psi_i(\theta)\}$ can be applied for both of the previous equations. For the new basis, you have

$$AF_0(\theta) = \sum_{i=1}^N B_i^0 \Psi_i(\theta), F(\theta) = \sum_{i=1}^N B_i \Psi_i(\theta) \quad (3-77)$$

$$B_i^0 = \langle AF_0(\theta), \Psi_i(\theta) \rangle, B_i = \langle F(\theta), \Psi_i(\theta) \rangle \quad (3-78)$$

I_i and A_i are finally found.

$$I_i = \sum_{j=1}^N B_j^0 c_i^{(j)}, A_i = \sum_{j=1}^N B_j c_i^{(j)} = j\beta(\delta d_i) I_i \quad (3-79)$$

A significant change of the initial pattern $AF_0(\theta)$ can be obtained by using the procedure iteratively. The final positions of the array elements are found in the last iteration, where the desired approximation of the pattern is obtained. A Chebyshev pattern $T_{12}(x)$ with $SLL = -25$ dB and $HPBW = 9.2^\circ$ for a 13-element array and the amplitudes are shown in Figure 3-13 and in Table 3-3.

Synthesis as an Optimization Procedure

Synthesis of an antenna array can be characterized as a nonlinear optimization procedure. With this procedure, one or more convenient real functions that take their optimum values are constructed. These functions fit several antenna characteristics. Such characteristics are the radiation pattern, the impedance, an index, the antenna coupling, the excitation, the geometry, the size, the loadings, and the current of the antenna elements.

Optimization methods make use of the values of the optimization function or look at the gradient of this function. Optimization functions, in some cases, are not explicit functions, but they are computed numerically.

Except for the previous procedures, there are methods based on random searches. These methods make use of a random number generator that helps to determine successive points. Moreover, the simulated annealing and the Genetic algorithms are global optimizers that have found applications in antenna synthesis.

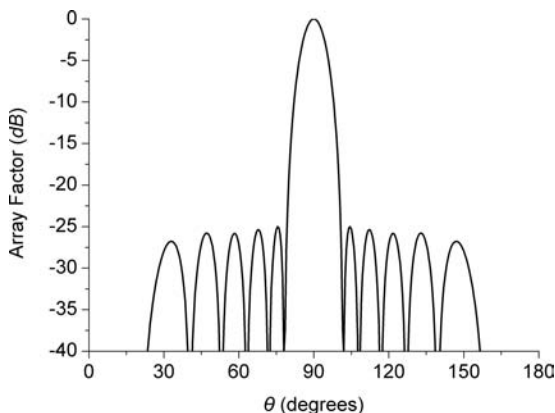


FIGURE 3-13 A $T_{12}(x)$ pattern with $SLL = -25$ dB and $HPBW = 9.2^\circ$ taken from an array with 13 elements

TABLE 3-3 Amplitudes and Positions of the Elements of the Array with the Pattern in Figure 3-13

Element Number	1/13	2/12	3/11	4/10	5/9	6/8	7
Amplitude	1.000	1.000	1.000	2.038	2.038	2.038	2.038
Position (λ)	± 3.006	± 2.436	± 2.065	± 1.525	± 0.952	± 0.467	0.000

Optimization of an Index In synthesis problems, the general purpose for a given array geometry is to derive the excitations without or under certain constraints. Most of the procedures that are commonly used suppose the existence of passive feed networks. These networks consist of suitable power combiners that concentrate all the element ports into one port. Modern arrays in corporate feed networks make use of *digital beamforming*. The pattern information is first digitized for each element and is then added together in a digital processor. The whole system, known as a *digital corporate feed*, produces the same pattern with the classical corporate structure. The element excitation at the feeder output can be found by pattern optimization. Casting certain array indices into an expression that is the ratio of Hermitian quadratic forms creates the procedure. An antenna index, I , such as directivity, gain, and so on, can be written as

$$I = \frac{[\tilde{u}]^* [A] [u]}{[\tilde{u}]^* [M] [u]} \quad (3-80)$$

$[\tilde{u}]^* = [u_1 \ u_2 \ \dots \ u_N]^*$ is the conjugate transpose of $[u]$. With $[u]$, you can represent the current or the voltage excitation vector of the array:

$$\left. \begin{aligned} [A] &= [\alpha_{ij}] \\ [M] &= [m_{ij}] \end{aligned} \right\} \quad (3-81)$$

$[A]$ and $[M]$ are Hermitian $N \times N$ square matrices and $[M]$ is a positive definite matrix. An optimization case is created by maximizing the index I subject to constrain another index I_1 , which equals a specified value.

$$I_1 = \frac{[\tilde{u}]^* [M_2] [u]}{[\tilde{u}]^* [M_3] [u]} = \gamma \quad (3-82)$$

Using the Lagrange multiplier and by setting the quantity L ,

$$L = \frac{[\tilde{u}]^* [A] [u]}{[\tilde{u}]^* [M] [u]} + \lambda \left\{ \frac{[\tilde{u}]^* [M_2] [u]}{[\tilde{u}]^* [M_3] [u]} - \gamma \right\} \quad (3-83)$$

stationary with respect to $[u]$ and λ you have

$$[u] = q [K]^{-1} [\tilde{B}]^* \quad (3-84)$$

q is a constant, and matrix $[K]$ is

$$[K] = [M] + p \{ \gamma [M_3] - [M_2] \} \quad (3-85)$$

The parameter p is found from the constraint equation (Eq. 3-84). You can find more details in the literature.¹³ If the parameter p is zero, optimization of I is without constraints.

Optimization by Simplex, Gradient, Simulated Annealing Methods, and Genetic Algorithms (GAs) The simplex method calculates the values of the optimization function at the vertices of a simplex. The term *simplex* means a body in multidimensional space. Based on the values of the optimization function, a smaller simplex is chosen within which an optimum should be created. The optimum depends on the initial simplex.¹⁴

The gradient or steepest-descent methods choose a starting point in the direction where the optimization function decreases most rapidly. By adopting a new point in that direction at a desired distance and then repeating the process, you achieve a minimum of the optimization function. Note, however, that the minimum is not always the global function. From an antenna engineering point-of-view, a suite solution is, in many cases, enough.

Simulated annealing combines local search with Monte Carlo techniques in analogy to the cooling processes in thermodynamics.¹⁵ *Annealing* refers to a process used to reveal the low temperature state of some material. The time spent at each temperature must be of sufficient length to allow thermal equilibrium to be realized. *Simulated annealing* means to simulate the annealing process using a Monte Carlo method where the global minimum of the objective function represents the low energy configuration.

In the simulated annealing process, parallelization techniques with the use of multiple CPUs are useful. Simulated annealing methods have been used in antenna design.¹⁶ The idea is to randomly change the array configuration (element position, amplitude, and phase). All or some of the antenna parameters are perturbed during the annealing process. The objective function can be the approximation of a desired pattern with another criterion in the sidelobe level and the antenna impedance in one or more frequencies.

Genetic algorithms are global optimizers in contrast to the local optimizer methods such as simplex and conjugate gradient methods. Genetic algorithms (GAs) can encode and improve complex structures using simple transformations. They are suitable for constrained optimization problems and are based on Darwin's principle¹⁷: survival of the fittest. The function to be optimized represents fitness.

The main concepts in GAs are selection, crossover, and mutation. Selection is the process of selecting the individuals to be mated. Selection can follow both probabilistic and deterministic rules based on the fitness of each individual. Crossover occurs with a probability P_c (usually 0.6–0.9) and involves the combination of the genetic information of the two parents. The crossover operator creates two new individuals with genetic characteristics from both parents. There is a probability of mutation for each new individual P_m (usually between 0.01–0.05). Mutation is a way of expanding the solution space to new unexplored domains.

GAs, in many cases due to algorithm-inherited complexity, require long computational times. To increase efficiency, an evolutionary algorithm based on the bird fly has been used recently. This algorithm is called *Particle Swarm Optimization (PSO)*.¹⁸ PSO is easy to implement and is used for single- and multi-objective optimization.

3.7 SMART ANTENNAS

Smart antennas are antenna arrays that are combined with signal processing in space and real time. They can be used in cellular and satellite mobile communications. Smart antennas increase spectrum efficiency and channel capacity. By multiplying beamsteering and electronically compensating distortion, they can extend the range of coverage. Smart antennas can provide a certain channel in a certain direction. They can also reduce propagation problems (multipath fading, co-channel interference, delay spread) and improve communication indices (bit error rate (BER) and outage probability, for example).

Smart antennas are characterized as adaptive arrays, intelligent antennas, spatial processing, digital beamforming antennas, among others.¹⁹ Two types of smart systems are available: *switchedbeam* and the *adaptive system*. The switched-beam antenna divides the communication area into micro-sectors. Each micro-sector contains a predetermined fixed-beam pattern. The adaptive systems dynamically alter the patterns to optimize communications performance. Adaptive array theory is based on optimization methods and on real-time response in a transient environment.

There are many special issues, books and specialized research papers in the area of smart antennas.

Smart antennas can be analyzed for different network topologies and mobility scenarios. Their geometries are realized with the appropriate feed networks and the algorithms for fast beamforming and direction of arrival. The cost of smart antennas continues to be the most critical issue.

3.8 ELEMENT PATTERN AND MUTUAL COUPLING

In the analysis and synthesis of antenna arrays, it was assumed that the characteristics of the elements were proportional to their excitations and that they were the same for similar elements and unchanged as the array was scanned. Actually, all of the currents and fields differ in magnitude, phase, and distribution from element to element. This happens because mutual coupling is involved in the behavior of the elements. In the literature, you will find several examples of a different pattern of an isolated element and the one in an array.^{3,7} An antenna array that shows more visually the importance of mutual coupling is the Yagi-Uda. In the Yagi-Uda array, due to the coupling, one of the elements (feeder) is enough to excite the parasitic ones (reflector, directors). Yagi-Uda produces easily directive patterns. Figure 3-14 presents the pattern of a four-element Yagi-Uda array.

In antenna array synthesis, the required distributions of the elements have to be derived. Different methods depend on the problem, such as the boundary value; the transmission line and the Poynting vector method are discussed in the literature. In the late 1960s, the integral

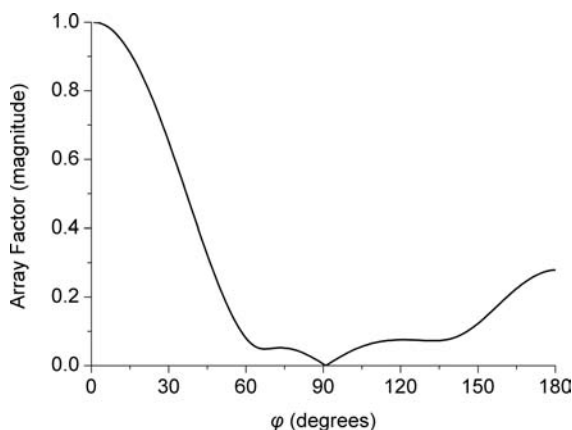


FIGURE 3-14 H-pattern of a four-dipole Yagi-Uda array. The distance between the elements is $d/\lambda = 0.30$, and the lengths are as follows: reflector = 0.50λ , feeder = 0.465λ , directors = 0.45λ , and dipole radius = 0.001λ .

equations with suitable numerical techniques, as the Method of Moments (MoM)²⁰ were given. The MoM reduces the integral equations to a system of simultaneous linear algebraic equations in terms of the unknown current or aperture distribution. There are computer codes, such as NEC and MININEC, used for evaluating the radiation characteristics of antennas.

Obviously, an antenna array needs the appropriate feed network to operate. There are several feeds, but the most usual are the series and the shunt feeds. In feed networks, impedance matching and the isolation between the outputs are critical functions. In the literature, you can find extended analysis of feed systems.^{1,7}

REFERENCES

1. C. A. Balanis, *Antenna Theory, Analysis and Design*, 3rd Ed. (New York: John Wiley & Sons Inc., 2005).
2. Y. T. Lo and S. W. Lee, *Antenna Handbook* (New York: Van Nostrand Reinhold, 1988).
3. J. N. Sahalos, *Orthogonal Methods for Array Synthesis, Theory and the ORAMA Computer Tool*, (New York: John Wiley & Sons Inc., 2006).
4. C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship Between Beam Width and Side-Lobe Level," *Proc. IRE*, vol. 34 (1946): 335–338.
5. S. A. Schelkunoff and H. T. Friis, *Antenna Theory and Practice*, (New York: John Wiley & Sons Inc., 1952).
6. H. J. Riblet, "Discussion on a Current Distribution for Broadside Arrays which Optimizes the Relationship Between Beam Width and Side-Lobe Level," *Proc. IRE*, vol. 35 (1947): 489–492.
7. J. N. Sahalos, "Antenna Arrays," *Wiley Encyclopedia of Telecommunications*, Ed. J.G. Proakis, vol. 1 (New York: John Wiley & Sons, Inc., 2004): 141–169.
8. G. Miaris, M. Chryssomalis, E. Vafiadis, and J. N. Sahalos, "A Unified Formulation for Chebyshev and Legendre Superdirective End-fire Array Design," *Archiv für Elektrotechnik*, vol. 78, 4 (1995): 271–280.
9. R. S. Elliot, "On Discretizing Continuous Aperture Distributions," *IEEE Trans. on Antennas & Propagation*, vol. AP-25, no. 5 (1977): 617–621.
10. R. C. Mailloux, *Phased Array Antenna Handbook* (Norwood, MA: Artech House, 2005).
11. A. T. Villeneuve, "Taylor Patterns for Discrete Arrays," *IEEE Trans. on Antennas & Propagation*, vol. AP-32, no. 10 (1984): 1089–1093.
12. E. T. Bayliss, "Design of Monopulse Antenna Difference Patterns with Low Sidelobes," *Bell System Tech. J.*, vol. 47 (1968): 623–650.
13. P. Zimourtopoulos and J. N. Sahalos, "On the Gain Maximization of the Dual Frequency and Direction Array Consisting of Wire Antennas," *IEEE Trans. on Antennas & Propagation*, vol. AP-33 (1985): 874–880.
14. J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Comp. Journal*, vol. 7 (1965): 308–313.
15. N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, "Equations of State Calculations by Fast Computing Machines," *Journal of Chemistry and Physics*, vol. 21 (1953): 1087–1091.
16. Z. Zaharis, E. Vafiadis, and J. N. Sahalos, "On the Design of a Dual-Band Base Station Wire Antenna," *IEEE Antennas and Propagation Magazine*, vol. 42, no. 6 (2000): 144–151.
17. C. Darwin, *On the Origin of Species* (London: John Murray, 1859).
18. J. Robinson and Y. Rahmat-Samii, "Particle Swarm Optimization in Electromagnetics," *IEEE Trans. on Antennas & Propagation*, vol. AP-52 (2004): 397–407.
19. M. Chryssomalis, "Smart Antennas," *IEEE Antennas & Propagation Magazine*, vol. 42, no. 3 (2000): 129–136.
20. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design* (New York: John Wiley and Sons, 1998).

P · A · R · T · 2

Types and Design Methods

Chapter 4

Dipoles and
Monopoles

Chen-To Tai*

The University of Michigan

Stuart A. Long

University of Houston

CONTENTS

4.1	INTRODUCTION.....	4-4
4.2	CYLINDRICAL DIPOLES.....	4-4
4.3	BICONICAL DIPOLES.....	4-11
4.4	FOLDED DIPOLES.....	4-14
4.5	SLEEVE DIPOLES.....	4-16
4.6	EFFECTIVE HEIGHT OF ANTENNAS.....	4-21
4.7	COUPLED ANTENNAS.....	4-23
4.8	MONOPOLE ANTENNAS.....	4-24

* Deceased

4.1 INTRODUCTION

Since the publication of the first edition of this handbook several sections in the original chapter on linear antennas have become outdated and have been deleted in this edition. The availability of computer programs¹ for finding the impedance and other characteristics of antennas, particularly linear antennas, makes parametric tabulation of limited use. Only some essential formulas and design data are therefore included in this chapter.

For the entire subject of linear antennas, the book by R. W. P. King² remains authoritative. Another book³ by the same author on the tables of antenna characteristics contains the most comprehensive data on the characteristics of cylindrical antennas. Calculations on circular-loop antennas and some simple arrays are also found there.

A section on the effective height of antennas is included in this chapter. The use of this parameter in describing the transmitting and receiving characteristics of linear antennas and other simple structures is discussed in detail. In addition, material on the general formulation of receiving antennas is included so that engineers can apply the formulation for design purposes or for estimation of the coupling effect between elements made up of both linear and other types of antennas.

Antennas in lossy media are of great current interest. Unfortunately, the subject cannot be covered in this chapter because of limited space. The book by King and Smith⁴ on antennas in matter can be consulted for this subject, particularly for linear antennas embedded in a lossy medium.

4.2 CYLINDRICAL DIPOLES

Impedance as a Function of Length and Diameter

The impedance characteristics of cylindrical antennas have been investigated by many writers. Theoretical work has mainly been confined to relatively thin antennas (length-to-diameter ratio greater than 15), and the effect of the junction connecting the antenna proper and the transmission line is usually not considered. Among various theories, the induced-emf method⁵ of computing the impedance of a cylindrical antenna based upon a sinusoidal distribution is still found to be very useful. The formula derived from this method is extremely simple. It is, however, valid only when the half length of a center-driven antenna is not much longer than a quarter wavelength. In practice, this is the most useful range. To eliminate unnecessary computations, the formula has been reduced to the following form:⁶

$$Z_i = R(k\ell) - j \left[120 \left(\ln \frac{\ell}{a} - 1 \right) \cot k\ell - X(k\ell) \right] \quad (4-1)$$

where

Z_i = input impedance, Ω , of a center-driven cylindrical antenna of total length 2ℓ and of radius a

$k\ell = 2\pi(\ell/\lambda)$ = electrical length, corresponding to ℓ , measured in radians

The functions $R(k\ell)$ and $X(k\ell)$ are tabulated in Table 4-1 and plotted in Figure 4-1 for the range $k\ell \leq \pi/2$. For calculation purposes, these two functions can be approximated to within 0.5Ω by the following simple third-order polynomials:

$$R(k\ell) = -0.4787 + 7.3246k\ell + 0.3963(k\ell)^2 + 15.6131(k\ell)^3$$

$$X(k\ell) = -0.4456 + 17.0082k\ell - 8.6793(k\ell)^2 + 9.6031(k\ell)^3$$

TABLE 4-1 Functions $R(k\ell)$ and $X(k\ell)$ Contained in the Formula of the Input Impedance of a Center-Driven Cylindrical Antenna

$k\ell$	$R(k\ell)$	$X(k\ell)$	$k\ell$	$R(k\ell)$	$X(k\ell)$
0	0	0	0.9	18.16	15.01
0.1	0.1506	1.010	1.0	23.07	17.59
0.2	0.7980	2.302	1.1	28.83	20.54
0.3	1.821	3.818	1.2	35.60	23.93
0.4	3.264	5.584	1.3	43.55	27.88
0.5	5.171	7.141	1.4	52.92	32.20
0.6	7.563	8.829	1.5	64.01	38.00
0.7	10.48	10.68	$\pi/2$	73.12	42.46
0.8	13.99	12.73			

When the length of the antenna is short compared with a wavelength but still large compared with its radius, the same formula reduces to

$$(Z_i)_{\text{short}} = 20(k\ell)^2 - j120(k\ell)^{-1} \left(\ln \frac{\ell}{a} - 1 \right) \quad (4-2)$$

For antennas of half length greater than a quarter wavelength, a number of refined theories provide formulas for the computation of the impedance function. None of them, however, is simple enough to be included here. As far as numerical computation is concerned, Schelkunoff's method⁷ is relatively simpler than Hallén's.² It should be emphasized that all these theories are formulated by using an idealized model in which the terminal condition is not considered.

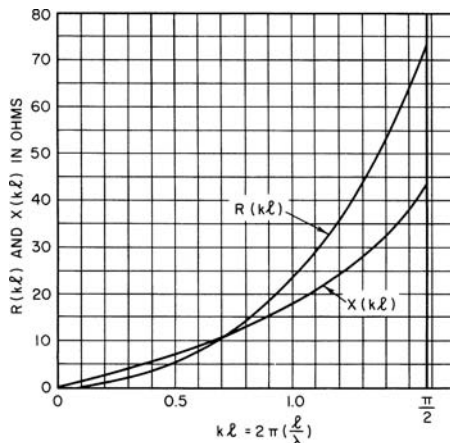


FIGURE 4-1 The functions $R(k\ell)$ and $X(k\ell)$

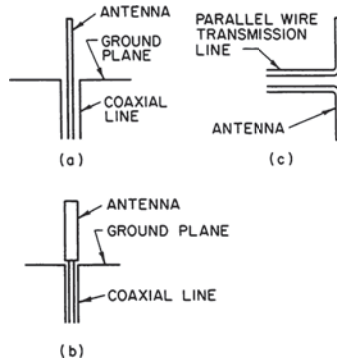


FIGURE 4-2 Driving an antenna by transmission lines

In practice, the antenna is always fed by a transmission line. The complete system may have the appearances shown in Figure 4-2. The effective terminal impedance of the line (often referred to as the *antenna impedance*) then depends not only upon the length and the diameter of the antenna but also upon the terminal condition. In cases *a* and *b*, the impedance would also be a function of the size of the ground plane. For a given terminal condition the variation of the impedance of a cylindrical antenna as a function of the length and the diameter of the antenna is best shown in the experimental work of Brown and Woodward.⁸ The data cover a wide range of values of the length-to-diameter ratio. Two useful sets of curves are reproduced in Figures 4-3 and 4-4. The impedance refers to a cylindrical antenna

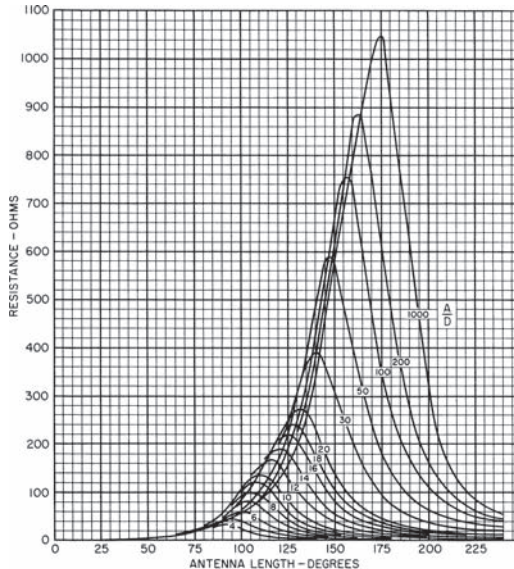


FIGURE 4-3 Antenna resistance versus antenna length A when a constant ratio of length to diameter A/D is maintained. The length and diameter are held constant while the frequency is changed.

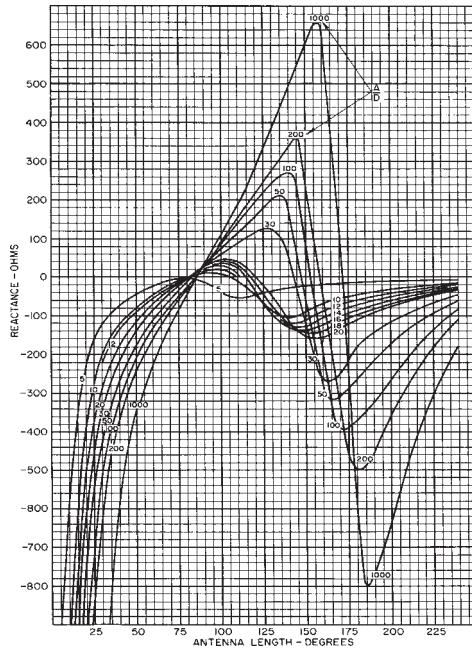


FIGURE 4-4 Reactance curves corresponding to the resistance curves of Figure 4-3

driven by a coaxial line through a large circular ground plane placed on the surface of the earth. The arrangement is similar to the one sketched in Figure 4-2a. The length and diameter of the antenna are measured in degrees; i.e., a length of one wavelength is equivalent to 360° . If the effects due to the terminal condition and finite-size ground plane are neglected, the impedance would correspond to one-half of the impedance of a center-driven antenna (Figure 4-2c). In using these data for design purposes, you must take into consideration the actual terminal condition as compared with the condition specified by these two authors. In particular, the maximum value of the resistance and the resonant length of the antenna may change considerably if the base capacitance is excessive.

Effect of Terminal Conditions

Many authors have attempted to determine the equivalent-circuit elements corresponding to different terminal conditions. Schelkunoff and Friis⁹ have introduced the concepts of *base capacitance* and *near-base capacitance* to explain the shift of the impedance curve as the terminal condition is changed. Similar interpretations have been given by King¹⁰ for a cylindrical antenna driven by a two-wire line or by a coaxial line and by Whinnery¹¹ for a biconical antenna driven by coaxial line. The importance of the terminal condition in affecting the input impedance of the antenna is shown in Figures 4-5 and 4-6. They are again reproduced from Brown and Woodward's paper. Because of the large variation of the effective terminal impedance of the line with changes in the geometry of the terminal junction, you must be cautious when using the theoretical results based upon isolated antennas. For junctions possessing simple geometry, the static method of Schelkunoff and Friis, King, and Whinnery can be applied to estimate the shunt capacitance of the junction.

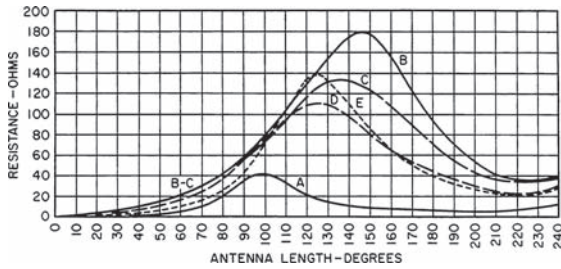


FIGURE 4-5 Resistance as a function of antenna length. A. The diameter D is 20.6° . Curve A: The arrangement shown in Figure 4-2a. Curve B: The arrangement of Figure 4-2a with the diameter of the outer conductor equal to 74° . The characteristic impedance of the transmission line is $77.0\ \Omega$. Curve C: The outer-conductor diameter is 49.5° , and the transmission line has a characteristic impedance of $52.5\ \Omega$. Curve D: The diameter of the outer conductor is 33° . The characteristic impedance is $28.3\ \Omega$. Curve E: This curve was obtained by tuning out the base reactance with an inductive reactance of $65.0\ \Omega$.

The latter then can be combined with the impedance of the antenna proper to evaluate the resultant impedance. For intricate junctions, accurate information can be obtained only by direct measurement.

Equivalent Radius of Noncircular Cross Sections

As far as the impedance characteristics and radiation pattern are concerned, a thin cylindrical antenna with a noncircular cross section behaves like a circular cylindrical antenna

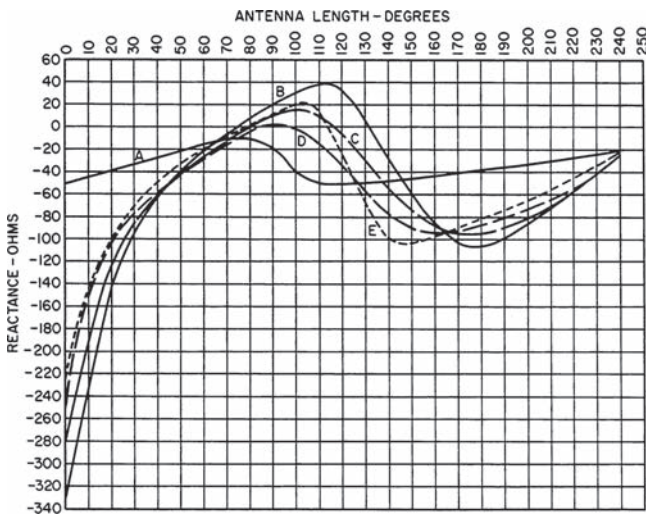


FIGURE 4-6 Reactance curves corresponding to the resistance curves of Figure 4-5

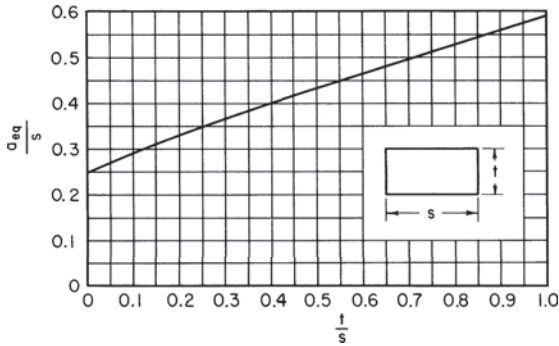


FIGURE 4-7 Equivalent radius a_{eq} of a rectangle as a function of the ratio of thickness t to width s

with an equivalent radius. In stating this characteristic, the terminal effect is, of course, not considered. The equivalent radius of many simply shaped cross sections can be found by the method of conformal mapping.¹² For an elliptical cross section, the following simple relation exists:

$$a_{eq} = \frac{1}{2}(a + b) \quad (4-3)$$

where

a = major half-axis of ellipse

b = minor half-axis of ellipse

For a rectangular cross section, the result is plotted in Figure 4-7. In the case of a strip, Eq. 4-3 and Figure 4-7 give the identical result of an equivalent radius equal to half the strip width. When the cross section has the form of a regular polygon, the result is tabulated in Table 4-2. The equivalent radius of two parallel cylinders of radius ρ_1 and ρ_2 separated by a distance d between the centers is given by¹³

$$\ln \rho_e = \frac{1}{(\rho_1 + \rho_2)^2} (\rho_1^2 \ln \rho_1 + \rho_2^2 \ln \rho_2 + 2\rho_1 \rho_2 \ln d) \quad (4-4)$$

Formulas for the equivalent radius of three cylinders and an angle strip are found in *Yagi-Uda Antenna* by S. Uda and Y. Mushiake.¹³

TABLE 4-2 Equivalent Radius of a Regular Polygon

n	3	4	5	6
a_{eq}/a	0.4214	0.5903	0.7563	0.9200

n = number of sides

a = radius of the outscribed circle

Patterns as a Function of Length and Diameter

In this subsection the radiation pattern of only center-driven cylindrical antennas is discussed. For base-driven antennas, the patterns depend very much upon the size of the ground plane. The subject will be discussed in Section 4.8.

The radiation pattern of a center-driven cylindrical antenna in general depends upon its length and thickness. The terminal condition that plays an important role in determining its impedance has a negligible effect on the pattern. For thin antennas, the calculated pattern obtained by assuming a sinusoidal current distribution is a good approximation of the actual pattern. Thus, with an assumed current distribution of the form

$$I(z) = I_0 \sin k(\ell - |z|) \quad +\ell \geq z \geq -\ell \quad (4-5)$$

the radiation field, expressed in a spherical coordinate system, is given by

$$E_\theta = \frac{j\eta I_0 e^{jkR}}{2\pi R} \left[\frac{\cos(k\ell \cos \theta) - \cos k\ell}{\sin \theta} \right] \quad (4-6)$$

where

$$\eta = (\mu/\epsilon)^{1/2} = 120\pi \Omega$$

θ = polar angle measured from axis of dipole, or z axis

The field pattern is obtained by evaluating the *magnitude* of the term contained in the brackets of Eq. 4-6. Some of the common patterns are shown in Figure 4-8. Comparing those patterns with the actual patterns of a thin cylindrical antenna obtained by measurement, reveals that the theoretical patterns based upon a sinusoidal current distribution do not produce the following characteristics found in actual measurements:

- The nulls between the lobes, except the *natural null* in the direction of the axis, are actually not zero.
- The phase of the field varies continuously from lobe to lobe instead of having a sudden jump of 180° between the adjacent lobes.
- The actual patterns vary slightly as a function of the radius of the antenna instead of being independent of the thickness.

Depending upon the particular applications, some of the fine details may require special attention. In most cases, the idealized patterns based upon a sinusoidal current distribution give us sufficient information for design purposes.

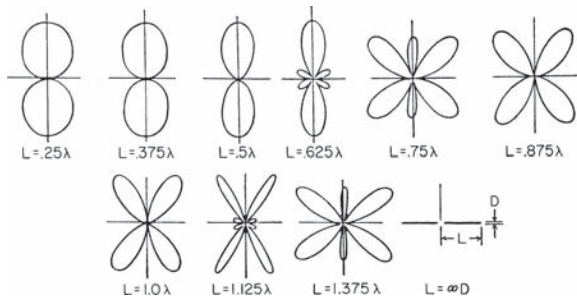


FIGURE 4-8 Radiation patterns of center-driven dipoles if sinusoidal current distribution is assumed

When the half length ℓ of the antenna is less than about one-tenth wavelength, Eq. 4-6 is well approximated by

$$E_{\theta} = \frac{j\eta I_0 (k\ell)^2 e^{jkR}}{4\pi R} \sin \theta \quad (4-7)$$

The figure-eight pattern resulting from the plot of the sine function is a characteristic not only of short cylindrical antennas but also of all small dipole-type antennas. Equations 4-6 and 4-7 are also commonly used to evaluate the directivity of linear antennas. The directivity is defined as

$$D = \frac{\text{maximum radiation intensity}}{\text{average radiation intensity}} \quad (4-8)$$

For a short dipole, D is equal to 1.5. The directivity of a half-wave dipole ($\ell = \lambda/4$) is equal to 1.64. The half-wave dipole is often used as a reference antenna to describe the gain of more directive antennas, particularly arrays made of dipoles.

4.3 BICONICAL DIPOLES

Impedance as a Function of Length and Cone Angle

When the angles of a symmetrical biconical antenna (Figure 4-9) are small, the input impedance of the antenna can be calculated using Schelkunoff's formula.¹⁴ Some sample curves are shown in Figure 4-10. While the biconical antenna is an excellent theoretical



FIGURE 4-9 A biconical dipole

model for studying the essential property of a dipole-type antenna, small-angle biconical antennas are seldom used in practice. Wide-angle biconical antennas or their derived types such as discones, however, are frequently used as broadband antennas. The broadband impedance characteristics occur when the angle of the cones, θ_0 of Figure 4-9, lies between 30 and 60°.

The exact value of θ_0 is not critical. Usually it is chosen so that the characteristic impedance of the biconical dipole matches as closely as possible the characteristic impedance of the line that feeds the antenna. The characteristic impedance of a biconical dipole as a function of the angle is plotted in Figure 4-11. For a conical monopole driven against an infinitely large ground plane, the characteristic impedance and the input impedance of the antenna are equal to half of the corresponding values of a dipole. Several formulas¹⁵ are available for computing the input impedance of wide-angle biconical antennas. Actual computation has been confined to a very few specific values of θ_0 .^{16,17} More complete information is available from the experimental data obtained by Brown and Woodward.¹⁸ Two curves are reproduced in Figures 4-12 and 4-13. The case corresponding to $\alpha = 0^\circ$ represents a cylindrical antenna having a diameter of 2.5 electrical degrees at a frequency of 500 MHz, since the feed point was kept fixed at that diameter.

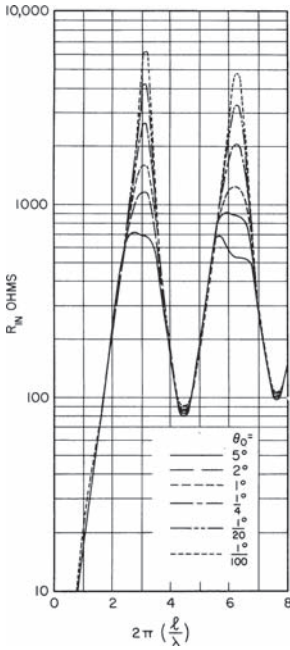


FIGURE 4-10a Input impedance of small-angle biconical antennas (resistance)

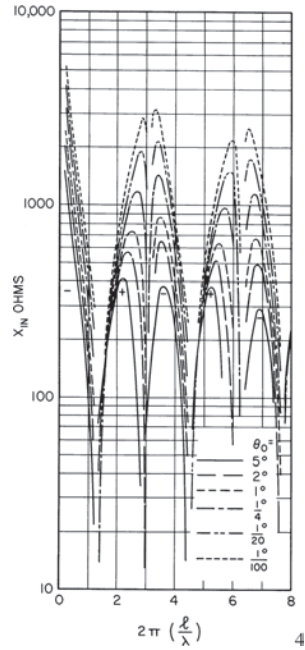


FIGURE 4-10b Input impedance of small-angle biconical antennas (reactance)

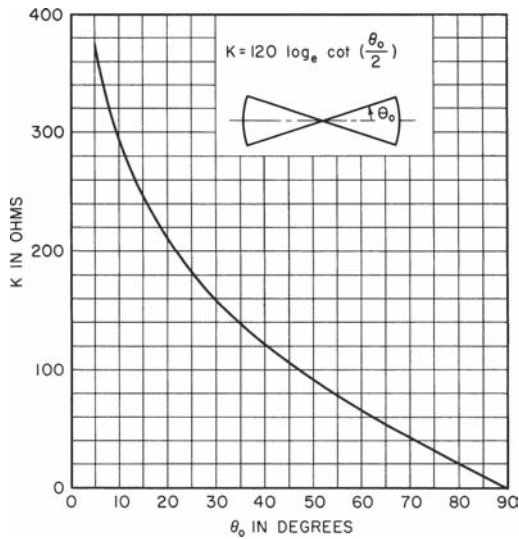


FIGURE 4-11 Characteristic impedance of a biconical dipole

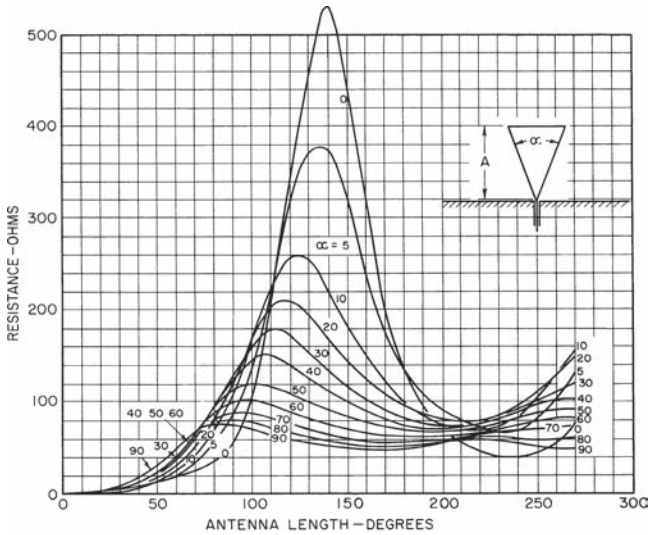


FIGURE 4-12 Measured resistance curves of the conical unipole versus length in electrical degrees for various flare angles

Patterns of the Biconical Dipole

The radiation patterns of biconical dipoles have been investigated theoretically by Papas and King.¹⁹ Figure 4-14 shows the patterns of a 60° -flare-angle ($\theta_0 = 30^\circ$) conical dipole for various values of ka , where $k = 2\pi/\lambda$ and a = half length of the dipole, which is the same as the ℓ used in Figure 4-9. Similar curves corresponding to different values of the flare angle have been obtained experimentally by Brown and Woodward.¹⁸

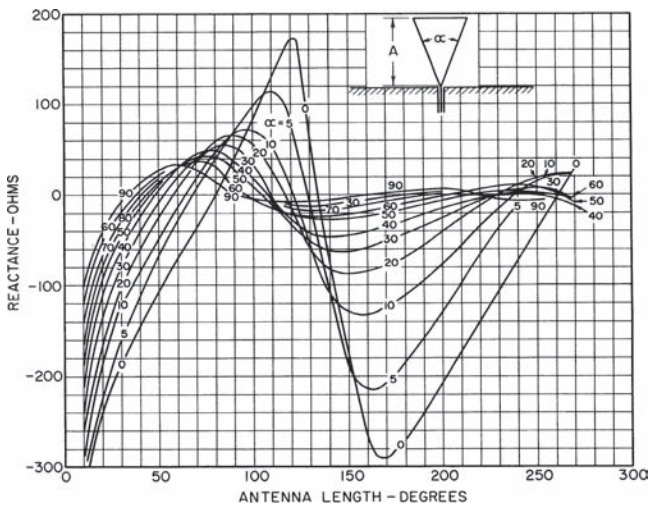


FIGURE 4-13 Measured reactance curves of the conical unipole versus length in electrical degrees for various flare angles

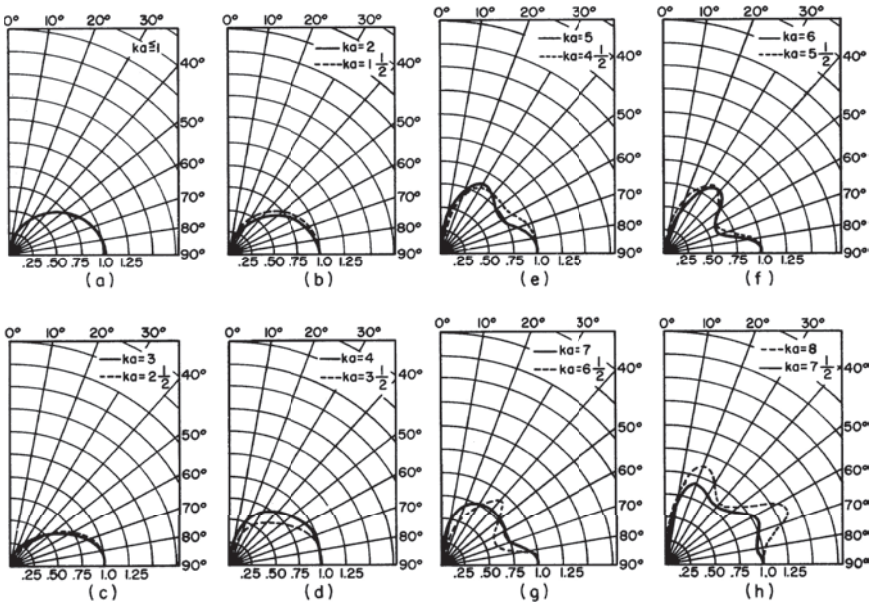


FIGURE 4-14 Plots of the absolute values of the far-zone electric field as a function of the zenithal angle θ for various values of ka and with a flare angle equal to 60° ($\theta_0 = 30^\circ$)

4.4 FOLDED DIPOLES

Equivalent Circuit of a Folded Dipole

A folded dipole is formed by joining two cylindrical dipoles at the ends and driving them by a pair of transmission lines at the center of one arm, as shown in Figure 4-15. The diameters of the two arms can be either identical or different. A simple analysis, based upon a quasi-static approach, of the operation of a folded dipole of arbitrary dimensions has been given by Uda and Mushiaki.¹³ According to their method, the excitation of a folded dipole can be considered as a superposition of two modes, as shown in Figure 4-16. The impedance of the

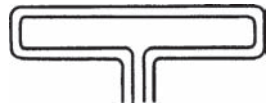


FIGURE 4-15 Folded dipole

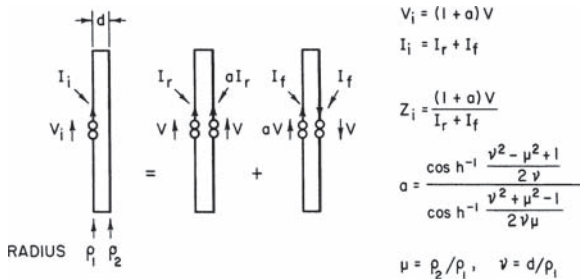


FIGURE 4-16 Decomposition of the folded dipole into two fundamental modes

symmetrical mode, characterized by two equal driving voltages, can be calculated by making use of the equivalent radius of two conductors as discussed in Section 4.2. The equivalence is shown in Figure 4-17. The impedance function Z_r is therefore the same as the impedance of a cylindrical dipole with an equivalent radius ρ_e given by

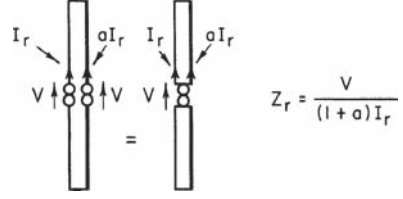


FIGURE 4-17 The equivalent representation of the symmetrical mode in computing Z_r

$$\ln \rho_e = \ln \rho_1 + \frac{1}{(1+\mu)^2} (\mu^2 \ln \mu + 2\mu \ln v) \quad (4-9)$$

where the various parameters are explained in Figure 4-16. The impedance of the asymmetrical mode, characterized by equal and opposite currents on the two arms, is the same as the shorted section of transmission line of length equal to ℓ ; that is,

$$Z_f = \frac{(1+a)V}{2I_f} = jZ_0 \tan k\ell \quad (4-10)$$

where Z_0 is the characteristic impedance of the two-wire line. Expressed in terms of Z_r and Z_f , the input impedance of a folded dipole is given by

$$Z = \frac{V_i}{I_i} = \frac{(1+a)V}{I_r + I_f} = \frac{2(1+a)^2 Z_r Z_f}{(1+a)^2 Z_r + 2Z_f} \quad (4-11)$$

An equivalent circuit based upon Eq. 4-11 is shown in Figure 4-18. For a folded dipole of length ℓ equal to $\lambda/4$, Z_f is very large compared with $(1+a)^2 Z_r$; hence

$$Z_{\lambda/4} = (1+a)^2 Z_r \quad (4-12)$$

Impedance Transformation as a Function of the Ratio of Conductor Sizes

The step-up impedance ratio $(1+a)^2$ as a function of μ and v has been calculated by Mushiake.²⁰ The diagram is reproduced in Figure 4-19 by using the formula for a given in

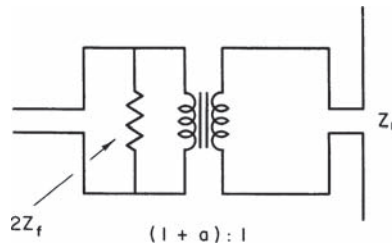


FIGURE 4-18 Equivalent circuit of a folded dipole

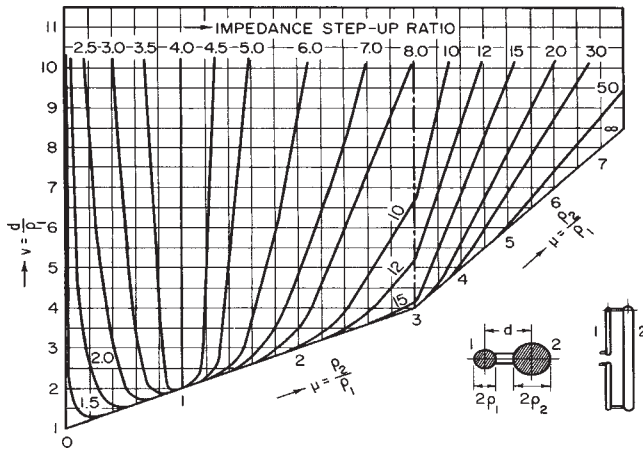


FIGURE 4-19 Step-up transformation chart for a folded dipole

Figure 4-16. When ρ_1 and ρ_2 are small compared with d , the value of a is given to a good approximation by

$$a = \frac{\ln(d/\rho_1)}{\ln(d/\rho_2)} \quad (4-13)$$

This formula was first derived by Guertler.²¹

Other presentations²² of the transformation ratio $(1 + a)^2$ in a logarithmic scale as a function of ρ_2/ρ_1 and d/ρ_1 are given in Figures 4-20 and 4-21.

4.5 SLEEVE DIPOLES

Equivalent Circuit of a Sleeve Dipole

The geometrical shape of a sleeve antenna, or a sleeve monopole, is shown in Figure 4-22a. If the image of the structure is included, then you have a sleeve dipole, as shown in Figure 4-22b. A sleeve dipole can therefore be considered as a doubly fed antenna in which the current is a relative maximum at the center of the dipole or at the base of the monopole. Antiresonances of the antenna impedance function take place when S is approximately equal to an odd multiple of a quarter wavelength or L is a multiple of a half wave. A special case of interest is that in which $L + S$ is equal to a quarter wavelength. Then the current distribution along the structure is approximately cosinusoidal. At resonance, the input resistance of the sleeve monopole or the sleeve dipole is approximately given by

$$R_s = R \left(\csc \frac{2\pi L}{\lambda} \right)^2 \quad (4-14)$$

where R denotes either the input resistance of a resonant quarter-wave monopole or that of a half-wave dipole. The sleeve in this case plays the role of an impedance transformer.

Wong and King²³ have shown experimentally that properly designed open-sleeve dipoles exhibit a broadband voltage-standing-wave-ratio (VSWR) response and unidirectional

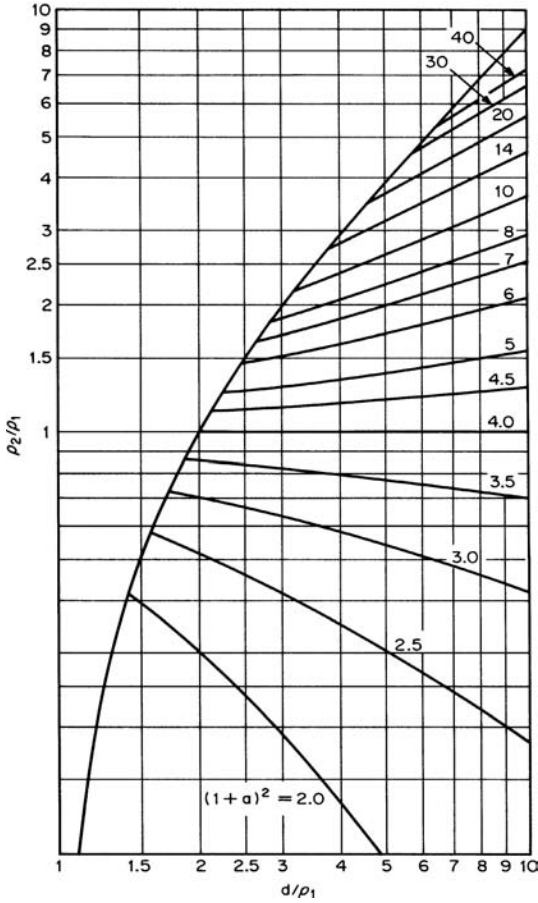


FIGURE 4-20 ρ_2/ρ_1 versus d/ρ_1 for various impedance transformation ratios (after Hansen²² © IEEE 1982)

radiation patterns over nearly an octave bandwidth when placed above a reflector. Figure 4-23 shows the VSWR response of open-sleeve dipoles for various dipole and sleeve diameters. Figure 4-24 shows the same response with sleeve spacing as the parameter.

Open Folded Sleeve Monopole

The work of Uda and Mushiake¹³ on folded dipoles can be extended to include a load at the unexcited arm, as shown schematically in Figure 4-25. The impedance of the loaded dipole is given by

$$Z_i = \frac{(Z_a + Z_L)Z_s + \left(\frac{a}{1+a}\right)^2 Z_L Z_a}{Z_s + Z_L + \frac{1}{(1+a)^2} Z_a} \quad (4-15)$$

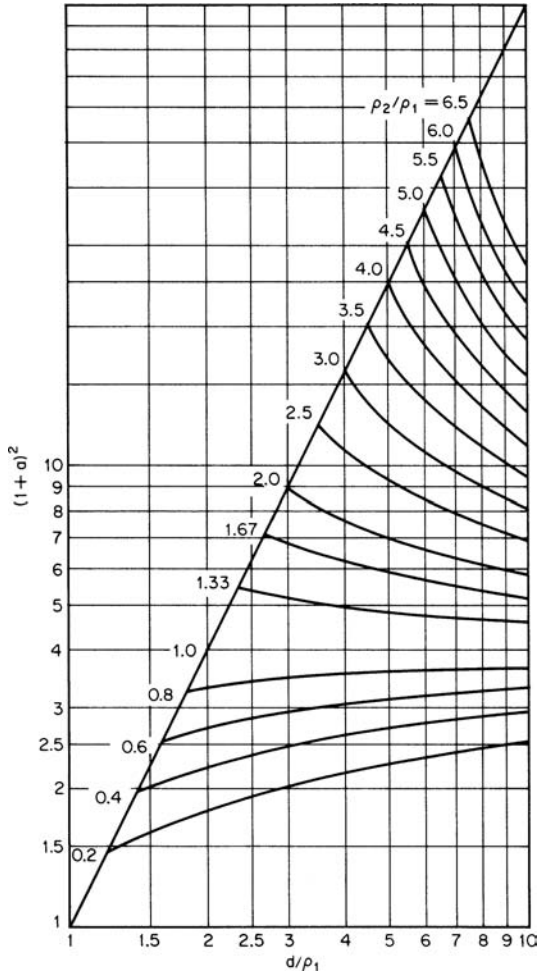


FIGURE 4-21 Impedance transformation ratio versus d/ρ_1 for various values of ρ_2/ρ_1 (after Hansen²² © IEEE 1982)

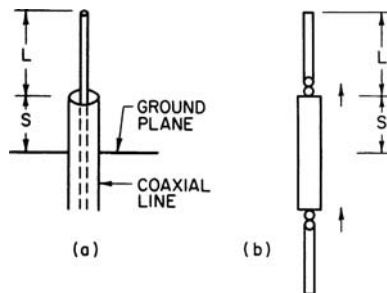


FIGURE 4-22 The sleeve antenna

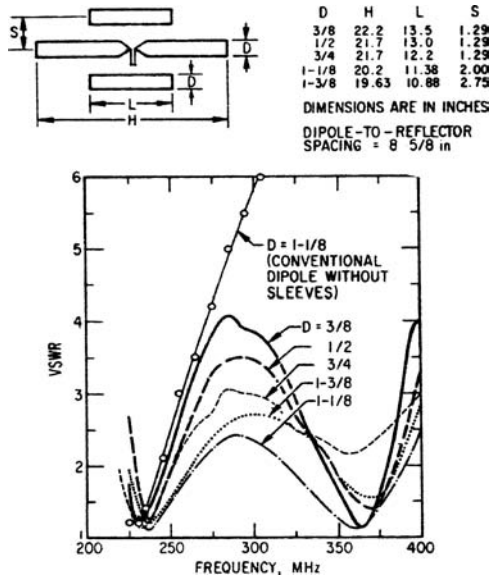


FIGURE 4-23 VSWR response of open-sleeve dipoles for various dipole and sleeve diameters

where

$$Z_a = jZ_c \tan \frac{2\pi \ell}{\lambda}$$

$$Z_c = \frac{Z_0}{2\pi} \left\{ \cos h^{-1} \left[\frac{1}{2v} (1 + \gamma^2 - \mu^2) \right] + \cos h^{-1} \left[\frac{1}{2v\mu} (-1 + \gamma^2 + \mu^2) \right] \right\}$$

Z_s = input impedance of the folded dipole when it is driven simultaneously by a common voltage at the base. It is approximately equal to that of a dipole with an equivalent radius ρ_e given by Eq. 4-4.

The parameter a in Eq. 4-15 and the parameters v and μ in the expression for the characteristic impedance Z_c of a transmission line made of wires of unequal radius are the same as those defined for Figure 4-16.

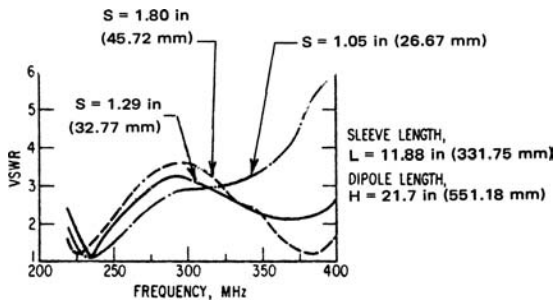


FIGURE 4-24 VSWR for a 3/4-in- (19-mm-) diameter open-sleeve dipole with sleeve spacing as the parameter

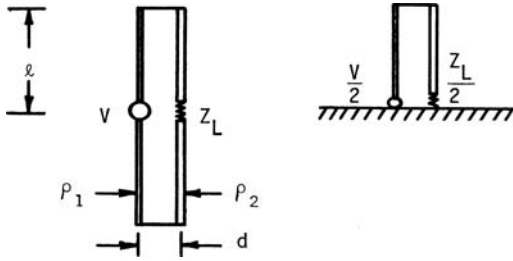


FIGURE 4-25 Loaded folded dipole and monopole

In the special case in which Z_L approaches infinity, Eq. 4-15 reduces to

$$Z_i = Z_s + \left(\frac{a}{1+a} \right)^2 Z_a \quad (4-16)$$

The structure then corresponds to an open folded dipole or monopole. Josephson²⁴ studied this structure both theoretically and experimentally. However, his analysis is correct only if the wires are of the same size. A sleeve version of an open folded monopole was also investigated by Josephson. The two structures are shown in Figures 4-26 and 4-27.

The open folded monopole with a displaced feed point is equivalent to a folded sleeve monopole. The input resistance at resonance of the open folded sleeve monopole is given approximately by

$$R_x = R_0 \left(\sin \frac{4\pi s}{\lambda} \right)^2 \left/ \left(\sin \frac{2\pi x}{\lambda} \right)^2 \right. \quad (4-17)$$

where R_0 denotes the resonant input resistance of the open folded monopole and R_x is the input resistance of the folded sleeve monopole. The precise value of R_0 depends on the radius of the wires and their separation. A typical value is about 10 Ω . The open folded sleeve monopoles made of arms with different sizes are useful models in designing aircraft trail antennas or fin-type antennas on vehicles as illustrated by Josephson.

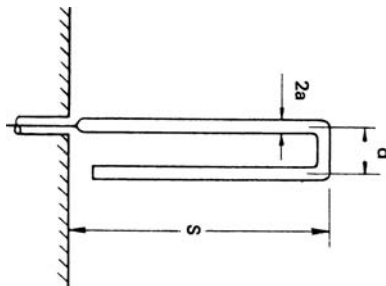


FIGURE 4-26 Open folded monopole

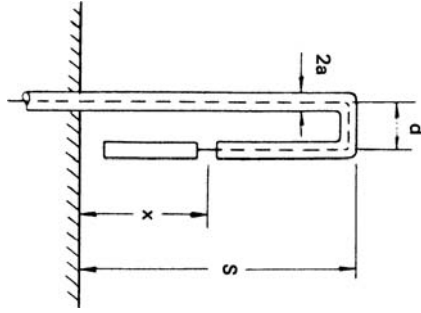


FIGURE 4-27 Open folded monopole with a displaced feed point

4.6 EFFECTIVE HEIGHT OF ANTENNAS

General Formula and Its Role in the Theory of Transmitting and Receiving Antennas

The radiation of any antenna can always be written in the form

$$\mathbf{E} = \frac{-jkZ_0 I_i e^{-jkR}}{4\pi R} \mathbf{h} \quad (4-18)$$

where

$$Z_0 = (\mu_0 / \epsilon_0)^{1/2}$$

$$k = 2\pi / \lambda$$

\mathbf{h} = effective height of antenna

I_i = input current to antenna

The effective height of an antenna was originally introduced by Sinclair.²⁵ It is related to the radiation vector defined by Schelkunoff to characterize the radiation field of an antenna, i.e.,

$$\mathbf{N}_t = I_i \mathbf{h} \quad (4-19)$$

where \mathbf{N}_t denotes the transversal part of Schelkunoff's radiation vector. The effective height is a very useful parameter in antenna engineering. For example, the open-circuit voltage of a receiving antenna can be expressed as

$$V_{op} = \mathbf{E}_i \cdot \mathbf{h} \quad (4-20)$$

where \mathbf{E}_i denotes the incident electric field. It is also an important parameter involved in the polarization-matching factor of a receiving antenna. In the theory of receiving antennas, the receiving cross section or the effective aperture is defined by²⁶

$$A = \frac{\lambda^2}{4\pi} Dpq \quad (4-21)$$

where

λ = operating wavelength

D = directivity of antenna

p = polarization-matching factor

$$= \frac{|\mathbf{h} \cdot \mathbf{E}_i|^2}{|\mathbf{h}|^2 |\mathbf{E}_i|^2}$$

q = impedance-matching factor

$$= 1 - \left| \frac{Z_L - Z_i^*}{Z_L + Z_i} \right|^2$$

Z_i = input impedance of antenna

Z_L = load impedance

The polarization-matching factor again involves the effective-height function. The effective-height function of a center-driven short dipole, if a linear current distribution along the dipole is assumed, is given by

$$\mathbf{h} = -\ell \sin \theta \hat{\theta} \quad (4-22)$$

where ℓ denotes the half length of the dipole that is assumed to be pointed in the vertical direction and θ denotes the polar angle measured between the axis of the dipole and the direction of observation or the direction of radiation. For a linearly polarized incident field making a skew angle α with the axis of the dipole, the polarization-matching factor is given by

$$p = \cos^2 \alpha \quad (4-23)$$

The effective-height functions for other simple antenna elements are listed below:

Antenna Type	Effective Height
Short dipole of length 2ℓ	$-\ell \sin \theta \hat{\theta}$
Half-wave dipole	$-\frac{\lambda}{\pi} \frac{\cos\left(\frac{\pi}{2} \cos \theta\right)}{\sin \theta} \hat{\theta}$
Small loop of radius a pointed in the z direction	$j \frac{2\pi^2 a^2}{\lambda} \sin \theta \hat{\phi}$
Half-wave folded dipole	$-\frac{2\lambda}{\pi} \frac{\cos\left(\frac{\pi}{2} \cos \theta\right)}{\sin \theta} \hat{\theta}$

4.7 COUPLED ANTENNAS

Circuit Relationships of Radiating Systems

When several antennas are coupled to each other, the input voltage and input currents to the antennas follow the same relationship as ordinary coupled circuits.²⁷ For a system of n antennas, the relationships are

$$V_i = \sum_{j=1}^n Z_{ij} I_j \quad i = 1, 2, \dots, n \quad (4-24)$$

where Z_{ii} is called the self-impedance of antenna i and Z_{ij} or Z_{ji} is called the mutual impedance between antenna i and antenna j . In the case of linear radiators, Carter's method, or the induced-emf method based upon sinusoidal current distribution, is the simplest one to use in determining the various Z 's. The method applies only to antennas shorter than a half wavelength. The self-impedance determined by this method is the same as that given by Eq. 4-1. The formulas for the mutual impedance of two parallel antennas of equal size are found in Carter's original paper or in Kraus's book.²⁸ Figure 4-28 shows the mutual impedance of two parallel half-wave antennas placed side by side. Figure 4-29 shows the mutual impedance of two parallel collinear half-wave antennas. Mutual impedances of two parallel antennas of unequal sizes have been investigated by several authors.²⁹⁻³¹ The induced-emf method has also been applied to crossed or skewed antennas^{32,33} to evaluate their mutual impedance. Refined calculations based upon Hallén's integral-equation technique are found in the works of Tai,³⁴ Bouwkamp,³⁵ and Uda and Mushiake.¹³ The last two authors also evaluated the self-impedance and mutual impedance of parallel antennas of unequal sizes, which ultimately applies to the design of Yagi-Uda arrays.

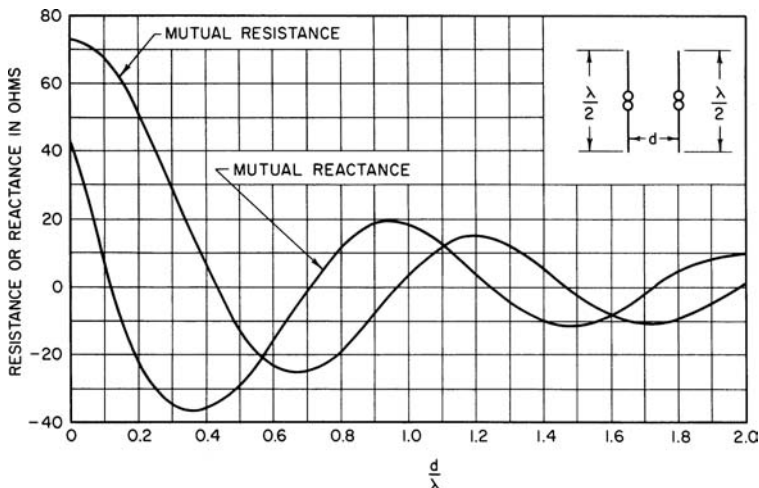


FIGURE 4-28 Mutual impedance between two parallel half-wave antennas placed side by side

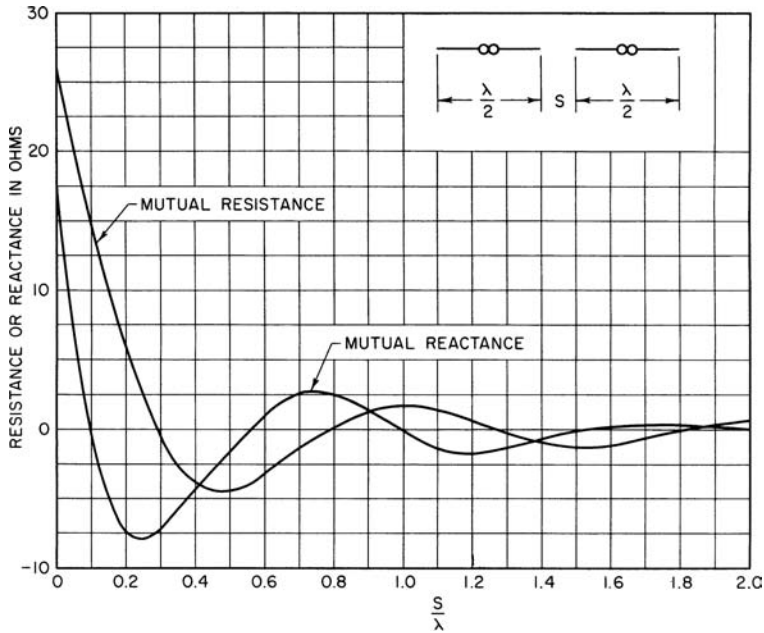


FIGURE 4-29 Mutual impedance between two collinear half-wave antennas

For dipoles separated by a distance that is large compared with a wavelength, the mutual impedance between the two dipoles can be calculated by using the asymptotic formula

$$Z_{12} = \frac{jkZ_0 e^{-jkR}}{4\pi R} (\mathbf{h}_1 \cdot \mathbf{h}_2) \quad (4-25)$$

where

$$k = 2\pi/\lambda$$

$$Z_0 = (\mu_0/\epsilon_0)^{1/2}$$

R = distance between centers of dipoles

$\mathbf{h}_1, \mathbf{h}_2$ = effective height of dipoles

The formula is quite accurate for half-wave dipoles with a separation barely greater than one wavelength.

4.8 MONOPOLE ANTENNAS

Relationship to Balanced Antennas

When a monopole is mounted on an ideally infinite ground plane, its impedance and radiation characteristics can be deduced from that of a dipole of twice its length in free space. For a base-driven monopole, its input impedance is equal to one-half that of the center-driven dipole, and the radiation pattern above the infinite ground plane is identical with the upper half of the radiation pattern of the corresponding dipole. When the ground plane is of finite size, the image theorem does not apply.

Several methods have been devised to investigate the characteristics of a monopole mounted on a finite-size ground plane. The first method is due to Bolljahn,³⁶ who considers the problem from the point of view of symmetrical components. The decomposition

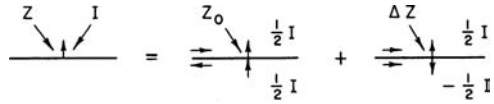


FIGURE 4-30a Monopole and finite-size ground plane and its decomposition into two modes of excitation

is shown in Figure 4-30a, in which the ground plane is assumed to be of the form of an infinitely thin conducting disk. For the symmetrical mode of excitation, the presence of the disk has no effect upon the radiation of the two elements. The problem is therefore the same as if the two elements were placed in free space. The antisymmetrical pair of current elements excites equal currents on the top and the bottom sides of the disk. This mode is responsible for the variation of the input impedance of the antenna as a function of the disk diameter. It is also responsible for the asymmetry of the resultant radiation pattern with respect to the ground plane. Bolljahn's original work was developed by assuming a short monopole on a disk. The entire analysis is found in Schelkunoff's book *Advanced Antenna Theory*.⁷ His study of the characteristics of large ground planes was later extended by Storer³⁷ to monopoles of arbitrary length.

Effect of Finite-Size Ground Plane on Impedance and Pattern

According to Storer, who used a variational method to formulate the problem, the change of the input impedance of a base-driven monopole erected upon a large circular ground plane can be written as

$$\Delta Z = Z - Z_0 = j \frac{60}{kd} e^{-jkd} \left| k \int_0^h \frac{I(z)}{I(0)} dz \right|^2 \quad (4-26)$$

where

Z_0 = impedance of monopole referred to an infinite ground plane, Ω

d = diameter of circular ground plane

$k = 2\pi/\lambda$

h = height of monopole

$I(z)$ = current-distribution function of monopole

$I(0)$ = base current or input current

The function $j(60/kd)e^{-jkd}$, which is independent of the current distribution, is plotted in Figure 4-30b. The real and the imaginary parts of the function are respectively equal to $(R - R_0) / \left| k \int_0^h \frac{I(z)}{I(0)} dz \right|^2$ and $(X - X_0) / \left| k \int_0^h \frac{I(z)}{I(0)} dz \right|^2$. For a quarter-wave monopole, if we assume $I(z) = I(0) \cos kz$, then

$$\left| k \int_0^h \frac{I(z)}{I(0)} dz \right| = 1$$

Thus, with a ground plane of a diameter greater than 10 wavelengths, you can see from Figure 4-30b that the variation of the resistance and of the reactance of a quarter-wave monopole is less than 1 Ω .

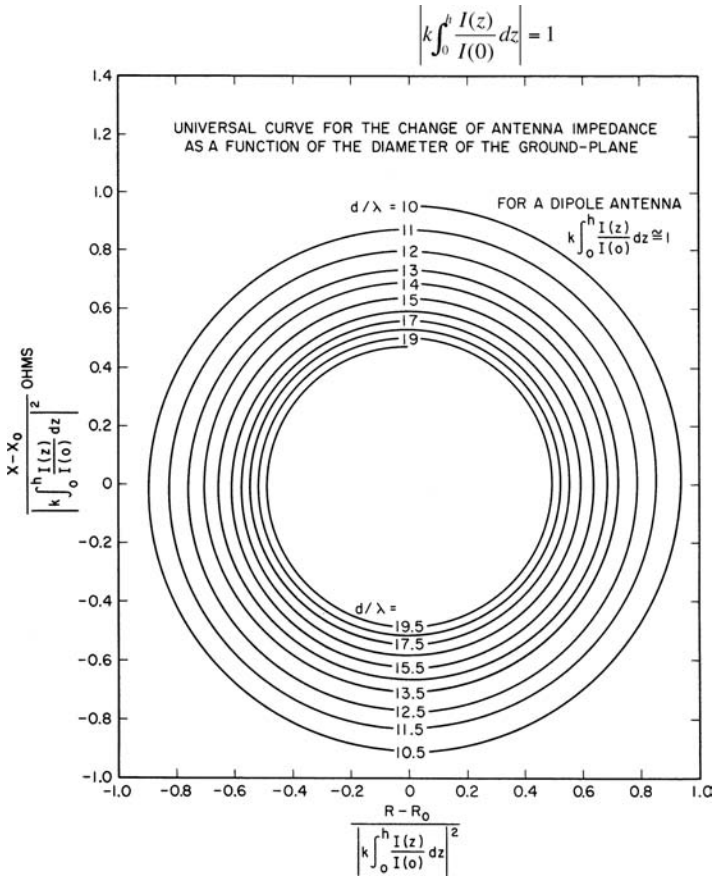


FIGURE 4-30b Universal curve for the change of antenna impedance as a function of the diameter of the ground plane

Radiation Pattern of a Monopole on a Circular Ground Plane*

While the effect of a ground plane upon the impedance of a monopole is not very significant, the radiation pattern is affected considerably. The pattern of such a composite antenna can be obtained quite accurately using the uniform geometrical theory of diffraction (GTD).^{38,39} For a short monopole of length h positioned at the center of a circular disk of radius a , as shown in Figure 4-31, the current on the monopole is assumed to be

$$I(z) = I_0 \frac{\sin k(h-z)}{\sin kh} \quad (4-27)$$

* The material contained in this subsection is condensed from a communication from Dr. Robert G. Kouyoumjian exclusively prepared for this handbook. The help of Dr. Kouyoumjian is gratefully acknowledged.

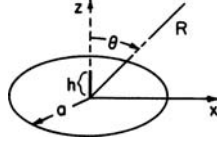


FIGURE 4-31 A monopole on a circular disk

If the point of observation is not near the vertical axis (θ is not close to 0 or π), the radiation field can then be expressed in the form

$$\mathbf{E} = (E_0 + E_{d1} + E_{d2}) \hat{\theta} \quad (4-28)$$

where E_0 is the geometrical optics field given by

$$E_0 = \begin{cases} \frac{jZ_0 I_0 e^{-jkR}}{2\pi R \sin kh} \left[\frac{\cos(kh \cos \theta) - \cos kh}{\sin \theta} \right], & 0 \leq \theta < \frac{\pi}{2} \\ 0, & \frac{\pi}{2} < \theta \leq \pi \end{cases} \quad (4-29)$$

and E_{d1} and E_{d2} represent, respectively, the singly and doubly diffracted field. They are given by

$$E_{d1} = -\frac{E_0 \left(a, \frac{\pi}{2}\right) a}{2\sqrt{2\pi ka \sin \theta}} \left\{ F \left[2ka \cos^2 \left(\frac{\pi}{4} + \frac{\theta}{2} \right) \right] \sec \left(\frac{\pi}{4} + \frac{\theta}{2} \right) e^{j[ka \sin \theta - (\pi/4)]} \mp \sec \left(\frac{\pi}{4} - \frac{\theta}{2} \right) e^{-j[ka \sin \theta - (\pi/4)]} \right\} \frac{e^{-jkR}}{R} \quad (4-30)$$

$$E_{d2} = -\frac{E_0 \left(a, \frac{\pi}{2}\right) a}{2\sqrt{2\pi ka \sin \theta}} \frac{e^{-j(2ka - \pi/4)}}{2\sqrt{\pi ka}} \cdot \left\{ F \left[2ka \cos^2 \left(\frac{\pi}{4} + \frac{\theta}{2} \right) \right] \sec \left(\frac{\pi}{4} + \frac{\theta}{2} \right) e^{j[ka \sin \theta - (\pi/4)]} \mp \sec \left(\frac{\pi}{4} - \frac{\theta}{2} \right) e^{-j[ka \sin \theta - (\pi/4)]} \right\} \frac{e^{-jkR}}{R} \quad (4-31)$$

where

$$F(x) = 2j\sqrt{x} e^{jx} \int_{\sqrt{x}}^{\infty} e^{-jt^2} dt \quad (4-32)$$

$$E_0 \left(a, \frac{\pi}{2}\right) = \frac{jZ_0 I_0 e^{-jka}}{2\pi a \sin kh} (1 - \cos kh) \quad (4-33)$$

In Eqs. 4-30 and 4-31, the upper sign associated with the secant functions is for $0 \leq \theta < \pi/2$, and the lower sign is for $\pi/2 < \theta \leq \pi$. Although the uniform GTD is a high-frequency method, which implies that $ka \gg 1$, in practice the result is valid even for ka as small as 2π . When the argument of the transition function $F(x)$ is greater than 10 or so, $F(x)$ is approximately equal to 1; then the sum of E_{d1} and E_{d2} yields

$$E_{d1} + E_{d2} = -E_0(a, \pi/2) \frac{a}{4} \sqrt{\frac{2}{\pi ka \sin \theta}} \left\{ \sec \left(\frac{\pi}{4} + \frac{\theta}{2} \right) e^{j[ka \sin \theta - (\pi/4)]} \right. \\ \left. \mp \sec \left(\frac{\pi}{4} - \frac{\theta}{2} \right) e^{-j[ka \sin \theta - (\pi/4)]} \left[1 + \frac{e^{-j[2ka - (\pi/4)]}}{2\sqrt{\pi ka}} \right] \right\} \frac{e^{-jkR}}{R} \quad (4-34)$$

In the axial region (θ close to 0 or π), the pattern can be found by means of an equivalent edge current.^{39,40} The total diffracted field in this case is given by

$$E_{d1} + E_{d2} = \mp jE_0 \left(a, \frac{\pi}{2} \right) \frac{a}{2} \sec \frac{\pi}{4} J_1(ka \sin \theta) \left[1 + \frac{e^{-j(2ka - \pi/4)}}{2\sqrt{\pi ka}} \right] \quad (4-35)$$

This approximately equals E_{d1} and E_{d2} given by Eqs. 4-30 and 4-31 at $ka \sin \theta \approx 5$ for $\theta \ll \pi/2$. Thus, the pattern calculated from Eq. 4-35 joins smoothly with that calculated by Eqs. 4-31 and 4-32 when $ka > 5$.

Figure 4-32 shows a typical pattern based on Eq. 4-28, the solid line, and Eq. 4-35, the dashed line. For practical purposes the term E_{d2} due to the doubly defracted rays is much smaller compared with E_{d1} except in the region where θ is close to $\pi/2$. By using a hybrid moment method jointly with GTD, it is possible to determine the input impedance of a monopole centered on a perfectly conducting circular disk.⁴¹

Monopole Mounted on the Edge of a Sheet

When a monopole is driven against an infinitely large conducting half sheet, the problem can be formulated conveniently with the aid of the dyadic Green's function pertaining to the half sheet.⁴² By transforming the resultant series into definite integrals, Sawaya, Ishizone, and Mushiaki⁴³ have been able to calculate the impedance of a monopole mounted on a half sheet in several orientations. Their results are shown in Figures 4-33 and 4-34.

These theoretical data can be used to estimate the impedance of monopoles mounted on large but finite conducting sheets. These authors also calculated the impedance of a notch antenna cut on a half sheet. The radiation patterns of short dipoles and small loops mounted on a half sheet are found in C. T. Tai's *Dyadic Green's Functions in Electromagnetic Theory*.⁴² Other work also has been reported by Pozar and Newman⁴⁴ and by Marin and Catedra⁴⁵ for the impedance and radiation patterns of monopoles located near the edges of finite half sheets or disks.

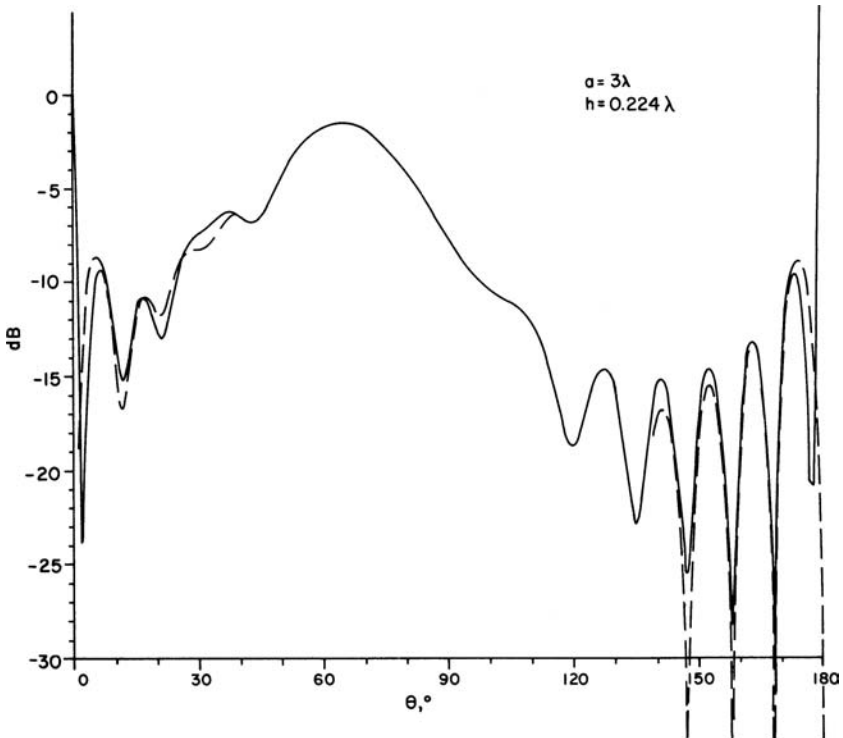


FIGURE 4-32 The radiation pattern of a monopole above a circular disk having a radius of three wavelengths

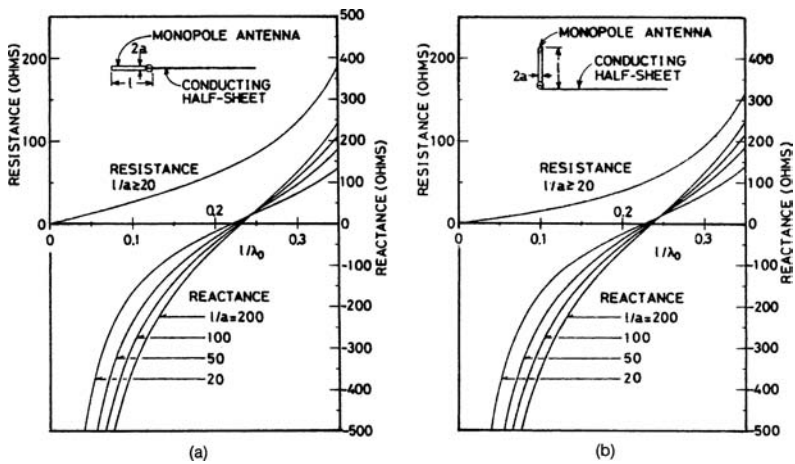


FIGURE 4-33 Impedance of a horizontal (a) and of a vertical (b) monopole antenna as a function of l/λ_0 (after Sawaya et al⁴³ © IEEE 1981)

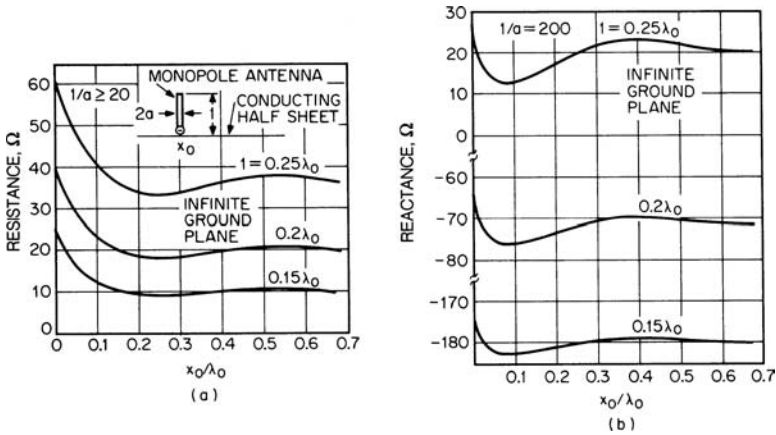


FIGURE 4-34 Impedance of vertical monopole antenna on a conducting half sheet as function of x_0/λ_0 , where x_0 is the distance from the edge of the half sheet to the base of the monopole. (a) Resistance (b) Reactance (after Sawaya et al⁴³ © IEEE 1981)

Monopole Mounted on a Conducting Box

Investigations by Bhattacharya et al.⁴⁶ and Chu et al.⁴⁷ have reported the impedance and radiation patterns of a monopole mounted at arbitrary locations on a conducting box that is comparable in size to a wavelength. Figure 4-35 shows the behavior of the admittance of a monopole as a function of frequency as the position of the monopole is moved from the center of the box face to one corner.⁴⁶

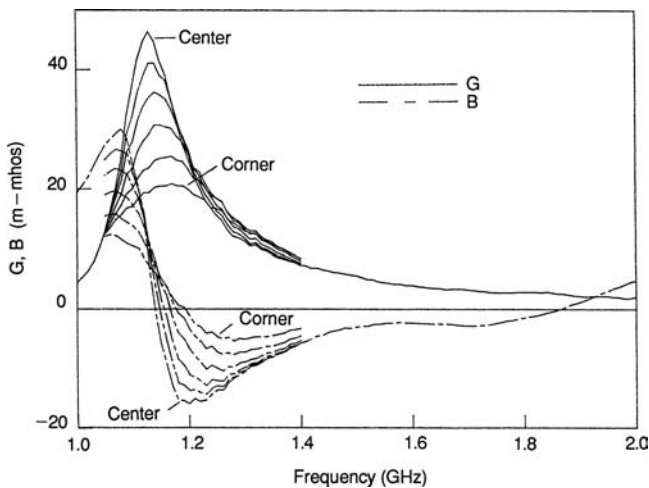


FIGURE 4-35 Admittance versus frequency of a 6-cm monopole mounted on a cubical conducting box with edge length 10 cm for various locations between the center and corner (after Bhattacharya et al⁴⁶ © IEEE 1987)

REFERENCES

1. G. J. Burke and A. J. Poggio, "Numerical Electromagnetic Code (NEC)," Tech. Doc. 116 (San Diego: Naval Ocean Systems Center, 1980).
2. R. W. P. King, *The Theory of Linear Antennas* (Cambridge: Harvard University Press, 1956).
3. R. W. P. King, *Tables of Antenna Characteristics* (New York: Plenum Press, 1971).
4. R. W. P. King and G. Smith, *Antennas in Matter* (Cambridge: The M.I.T. Press, 1981).
5. P. S. Carter, "Circuit Relations in Radiating Systems and Applications to Antenna Problems," *IRE Proc.*, vol. 20 (1932): 1004-1041.
6. R. S. Elliot, *Antenna Theory and Design* (Englewood Cliffs: Prentice Hall, Inc., 1981): 301-302.
7. S. A. Schelkunoff, *Advanced Antenna Theory* (New York: John Wiley & Sons, Inc., 1952).
8. George H. Brown and O. M. Woodward, Jr., "Experimentally Determined Impedance Characteristics of Cylindrical Antennas," *IRE Proc.*, vol. 33 (1945): 257-262.
9. S. A. Schelkunoff and H. T. Friis, *Antennas: Theory and Practice*, sec. 13.22 (New York: John Wiley & Sons, Inc., 1952): 445-448.
10. R. W. P. King, "Antennas and Open-Wire Lines, Part I, Theory and Summary of Measurements," *J. Appl. Phys.*, vol. 20 (1949): 832-850; "The End Correction for a Coaxial Line When Driving an Antenna over a Ground Screen," *IRE Trans. Antennas Propagat.*, vol. AP-3, no. 2 (April 1955): 66.
11. John R. Whinnery, "The Effect of Input Configuration to Antenna Impedance," *J. Appl. Phys.*, vol. 21 (1950): 945-956.
12. Y. T. Lo, "A Note on the Cylindrical Antenna of Noncircular Cross-Section," *J. Appl. Phys.*, vol. 24 (1953): 1338-1339.
13. S. Uda and Y. Mushiake, *Yagi-Uda Antenna* (Tokyo: Maruzen Co., Ltd., 1954): 19.
14. S. A. Schelkunoff, "Theory of Antennas of Arbitrary Size and Shape," *IRE Proc.*, vol. 29 (September 1941): 493; see also Ref. 7, Chap. 2.
15. Ref. 7, Chap. 2.
16. C. H. Papas and R. W. P. King, "Input Impedance of Wide-Angle Conical Antennas Fed by a Coaxial Line," *IRE Proc.*, vol. 37 (November 1949): 1269.
17. C. T. Tai, "Application of a Variational Principle to Biconical Antennas," *J. Appl. Phys.*, vol. 20 (November 1949): 1076.
18. G. H. Brown and O. M. Woodward, Jr., "Experimentally Determined Radiation Characteristics of Conical and Triangular Antennas," *RCA Rev.*, vol. 13, no. 4 (December 1952): 425.
19. C. H. Papas and R. W. P. King, "Radiation from Wide-Angle Conical Antenna Fed by a Coaxial Line," *IRE Proc.*, vol. 39 (January 1951): 49.
20. Y. Mushiake, "An Exact Step-Up Impedance Ratio Chart of a Folded Antenna," *IRE Trans. Antennas Propagat.*, vol. AP-3, no. 4 (October 1954): 163.
21. R. Guertler, "Impedance Transformation in Folded Dipoles," *J. Brit. IRE*, vol. 9 (September 1949): 344.
22. R. C. Hansen, "Folded and T-Match Dipole Transformation Ratio," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 1 (January 1982).
23. J. L. Wong and H. E. King, "An Experimental Study of a Balun-Fed Open Sleeve Dipole in Front of a Metallic Reflector," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (March 1972): 201. See also "Design Variations and Performance Characteristics of the Open-Sleeve Dipole," Aersp. Rep. TR-0073 (3404)-2 (Los Angeles: Electronics Research Laboratory, The Aerospace Corporation).
24. Bengt Josephson, "The Quarter-Wave Dipole," *IRE Wescon Conv. Rec.*, part I (August 1957): 77.
25. George Sinclair, "The Transmission and Reception of Elliptically Polarized Waves," *IRE Proc.*, vol. 38 (1950): 148.
26. C. T. Tai, "On the Definition of the Effective Aperture of Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-9 (1961): 224.

27. P. S. Carter, "Circuit Relations in Radiating Systems and Applications to Antenna Problems," *IRE Proc.*, vol. 20 (1932): 1004.
28. J. D. Kraus, *Antennas* (New York: McGraw-Hill, 1950).
29. C. R. Cox, "Mutual Impedance between Vertical Antennas of Unequal Heights," *IRE Proc.*, vol. 35 (November 1947): 1367.
30. G. Barzilai, "Mutual Impedance of Parallel Aerials," *Wireless Eng.*, vol. 25 (November 1948): 343.
31. R. G. Medhurst and S. D. Pool, "Comments on Mutual Impedance of Parallel Aerials," *Wireless Eng.*, vol. 28 (February 1951): 67.
32. L. Lewin, "Mutual Impedance of Wire Aerials," *Wireless Eng.*, vol. 28 (December 1951): 352.
33. R. G. Medhurst, "Dipole Aerials in Close Proximity," *Wireless Eng.*, vol. 28 (December 1951): 356.
34. C. T. Tai, "Coupled Antennas," *IRE Proc.*, vol. 36 (April 1948): 487.
35. C. J. Bouwkamp, "On the Theory of Coupled Antennas," *Philips Res. Rep.*, vol. 3 (June 1948): 213.
36. J. T. Bolljahn, "Antennas near Conducting Sheets of Finite Size," *Univ. California Dept. Eng. Rep.* 162 (December 1949).
37. J. E. Storer, "The Impedance of an Antenna over a Large Circular Screen," *J. Appl. Phys.*, vol. 12 (1951): 1058.
38. R. G. Kouyoumjian and P. H. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *IEEE Proc.*, vol. 62 (1974): 1448–1461.
39. R. G. Kouyoumjian, "The Geometrical Theory of Diffraction and Its Application," in R. Mittra (ed.), *Topics in Applied Physics, vol. 3: Numerical and Asymptotic Techniques in Electromagnetics* (Berlin: Springer-Verlag OHG, 1975): 204.
40. E. F. Knott and T. B. A. Senior, "Comparison of Three High-Frequency Diffraction Techniques," *IEEE Proc.*, vol. 62 (1974): 1468–1474.
41. G. A. Thiele and T. H. Newhouse, "A Hybrid Technique for Combining Moments with the Geometrical Theory of Diffraction," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (1975): 62–69.
42. C. T. Tai, *Dyadic Green's Functions in Electromagnetic Theory* (Scranton: INTEXT, 1971).
43. K. Sawaya, T. Ishizone, and Y. Mushiake, "A Simplified Expression for the Dyadic Green's Function for a Conducting Halfsheet," *IEEE Trans. Antennas Propagat.*, vol. AP-29 (September 1981): 749.
44. D. M. Pozar and E. H. Newman, "Analysis of a Monopole Mounted Near an Edge or a Vertex," *IEEE Trans. Antennas Propagat.*, vol. AP-30 (May 1982): 401–408.
45. M. Marin and M. F. Catedra, "A Study of a Monopole Arbitrarily Located on a Disk Using Hybrid MM/GTD Technique," *IEEE Trans. Antennas Propagat.*, vol. AP-35 (March 1987): 287–292.
46. S. Bhattacharya, S. A. Long, and D. R. Wilton, "The Input Impedance of a Monopole Antenna Mounted on a Cubical Conducting Box," *IEEE Trans. Antennas Propagat.*, vol. AP-35 (July 1987): 756–762.
47. A. W. C. Chu, S. A. Long, and D. R. Wilton, "The Radiation Pattern of a Monopole Antenna Attached to a Conducting Box," *IEEE Trans. Antennas Propagat.*, vol. AP-38 (December 1990): 1907–1912.

Chapter 5

Loop Antennas

Glenn S. Smith

Georgia Institute of Technology

CONTENTS

5.1 INTRODUCTION.....	5-2
5.2 ELECTRICALLY SMALL LOOPS.....	5-2
5.3 ELECTRICALLY LARGE LOOPS.....	5-9
5.4 SHIELDED-LOOP ANTENNA.....	5-18
5.5 PULSE-EXCITED CIRCULAR-LOOP ANTENNA	5-20
5.6 ADDITIONAL TOPICS	5-22

5.1 INTRODUCTION

The single-turn loop antenna is a metallic conductor bent into the shape of a closed curve, such as a circle or a square, with a gap in the conductor to form the terminals. A multiturn loop or coil is a series connection of overlaying turns. The loop is one of the primary antenna structures; its use as a receiving antenna dates back to the early experiments of Hertz on the propagation of electromagnetic waves.¹

The discussion of loop antennas is conveniently divided according to electrical size. Electrically small loops, those whose total conductor length is small compared with the wavelength in free space, are the most frequently encountered in practice. For example, they are commonly used as receiving antennas with portable radios, as directional antennas for radio-wave navigation, and as probes with field-strength meters. Electrically larger loops, particularly those near resonant size (circumference of loop/wavelength ≈ 1), are used mainly as elements in directional arrays.

The following symbols are used throughout the chapter:

λ = wavelength in free space at the frequency $f = \omega/2\pi$, when the complex harmonic time-dependence $\exp(j\omega t)$ is assumed

$\beta = 2\pi/\lambda$ = propagation constant in free space

$\zeta = \sqrt{\mu_0/\epsilon_0}$ = wave impedance of free space ($\approx 377 \Omega$)

b = mean radius of a circular loop or mean side length of a square loop

a = radius of loop conductor (All results presented are for thin-wire loops, $a/b \ll 1$.)

A = area of loop

N = number of turns

ℓ_c = length of solenoidal coil

5.2 ELECTRICALLY SMALL LOOPS

The axial current distribution in an electrically small loop is assumed to be uniform; that is, the current has the same value I_0 at any point along the conductor. For single-turn loops and multiturn loops that are single-layer solenoidal coils, measurements suggest that this is a good assumption provided the total length of the conductor ($N \times$ circumference) is small compared with the wavelength in free space, typically $\leq 0.1\lambda$, and the length-to-diameter ratio for the solenoidal coil is greater than about 3 ($\ell_c/2b \geq 3.0$).² With a uniform current assumed, the electrically small loop antenna is simply analyzed as a radiating inductor.³

Transmitting Loop

The electromagnetic field of an electrically small loop antenna is the same as that of a magnetic dipole with moment $m = I_0 NA$:

$$E_\phi = \frac{\zeta \beta^2 m}{4\pi r} \left(1 - \frac{j}{\beta r}\right) e^{-j\beta r} \sin \theta \quad (5-1)$$

$$B_\theta = \frac{-\mu_0 \beta^2 m}{4\pi r} \left(1 - \frac{j}{\beta r} - \frac{1}{\beta^2 r^2}\right) e^{-j\beta r} \sin \theta \quad (5-2)$$

$$B_r = \frac{\mu_0 \beta^2 m}{2\pi r} \left(\frac{j}{\beta r} + \frac{1}{\beta^2 r^2}\right) e^{-j\beta r} \cos \theta \quad (5-3)$$

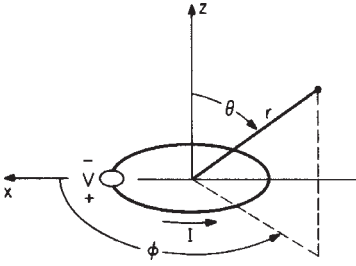


FIGURE 5-1 Loop antenna and accompanying spherical coordinate system

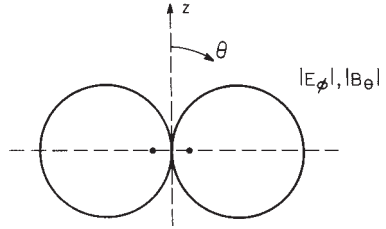


FIGURE 5-2 Far-zone vertical-plane field pattern of an electrically small loop

in which the plane of the loop is normal to the polar axis of the spherical coordinate system (r, θ, ϕ) centered at the loop, as shown in Figure 5-1. In the far zone of the loop ($\lim \beta r \rightarrow \infty$), only the leading terms in Eqs. 5-1 and 5-2 are significant, and the field pattern for both E_ϕ and B_θ in the vertical plane is the simple figure eight shown in Figure 5-2.

The driving-point voltage and current are related through the input impedance of the loop, $V = ZI_0$. For electrically small loops, the impedance is the series combination of the reactance of the external inductance L^e with the radiation resistance R^r and the internal impedance of the conductor $Z^i = R^i + j\omega L^i$:

$$Z = R + j\omega L = R^r + Z^i + j\omega L^e = R^r + R^i + j\omega(L^e + L^i) \quad (5-4)$$

The internal resistance R^i accounts for ohmic loss. In the equivalent circuit for the small loop, a lumped capacitance C is sometimes placed in parallel with Z to account for the distributed capacitance between the sides of a single turn and between the turns of a solenoid, as shown in Figure 5-3. Note that a loop with a truly uniform current distribution would have no capacitance, since from the equation of continuity there would be no charge along the conductor of the loop.

The radiation resistance of the small loop is proportional to the square of the product of the area and the number of turns:

$$R^r = \frac{\zeta}{6\pi} \beta^4 (NA)^2 \quad (5-5)$$

For single-turn loops and solenoidal coils whose turns are not too closely spaced, the internal impedance is approximately

$$Z^i = z^i \times \text{total length of conductor} \quad (5-6)$$

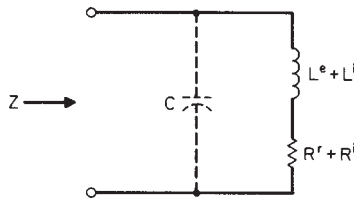


FIGURE 5-3 Equivalent circuit for input impedance Z of an electrically small loop

in which z^i is the internal impedance per unit length of a straight conductor with the same cross section as the loop conductor.⁴ If the turns of the coil are closely spaced, the proximity effect must also be included in determining Z^i .⁵

The external inductance is determined from one of the many formulas available for the inductance of coils:⁶ For a single-turn circular loop

$$L^e = \mu_0 b [\ln(8b/a) - 2] \quad (5-7)$$

and for a single-turn square loop

$$L^e = \frac{2\mu_0 b}{\pi} [\ln(b/a) - 0.774] \quad (5-8)$$

The external inductance of a tightly wound single-layer solenoidal coil of N turns, length ℓ_c , and radius b is often approximated by Lorenz's formula for the inductance of a circumferentially directed current sheet.⁶ Numerical results from this formula can be put in a form convenient for application:

$$L^e = K \mu_0 N^2 A / \ell_c \quad (5-9)$$

in which the factor K , known as Nagaoka's constant, is shown as a function of the ratio $\ell_c/2b$ (length of the coil to the diameter) in Figure 5-4. Note that, for a long coil ($\ell_c/2b \gg 1$), $K \approx 1$. The use of Eq. 5-9 assumes that the turns of the coil are so closely spaced that the winding pitch and insulation on the conductors can be ignored; if highly accurate calculations of L^e are necessary, corrections for these factors are available in the literature.⁶

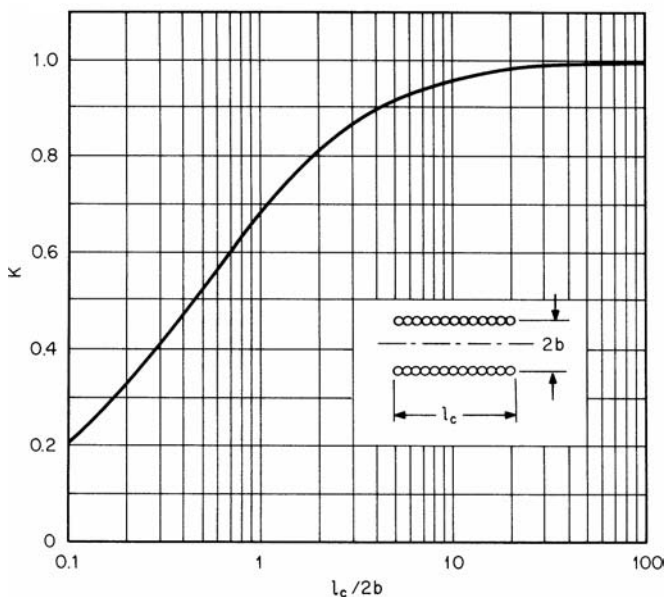


FIGURE 5-4 Nagaoka's constant K for a solenoidal coil as a function of the coil length to the diameter, $\ell_c/2b$

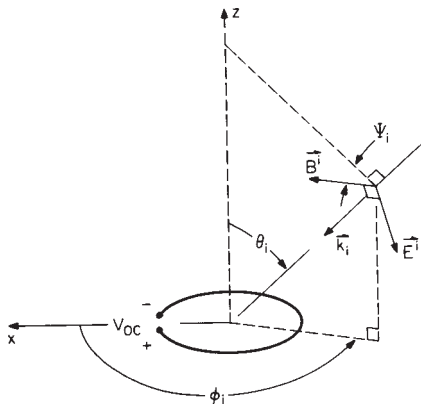


FIGURE 5-5 Plane-wave field incident on receiving loop

The radiation efficiency of the electrically small transmitting loop antenna is

$$\eta = \frac{\text{Time-Average Power Radiated}}{\text{Time-Average Power Supplied}} = \frac{R^r}{R^r + R^i} \quad (5-10)$$

The ohmic resistance R^i is often comparable to or larger than the radiation resistance R^r , so the radiation efficiency can be low.⁵ It can be increased by decreasing the ohmic resistance of the loop, i.e., by using a conductor of lower resistivity or larger radius.

Receiving Loop

When the electrically small loop is used as a receiving antenna, the voltage developed at its open-circuited terminals V_{OC} is proportional to the component of the incident magnetic flux density normal to the plane of the loop B_z^i :

$$V_{OC} = j\omega NAB_z^i \quad (5-11)$$

in which the incident field is assumed to be uniform over the area of the loop. This simple relation between V_{OC} and B_z^i makes the small loop useful as a probe for measuring the magnetic flux density. If a relation between the incident electric and magnetic fields at the center of the loop is known, V_{OC} can be expressed in terms of the magnitude of the incident electric field E^i and an effective height h_e . This is the case for an incident linearly polarized plane wave with the wave vector \mathbf{k}_i and the orientation shown in Figure 5-5:

$$V_{OC} = j\omega NAB^i \cos \psi_i \sin \theta_i = h_e(\psi_i, \theta_i)E^i \quad (5-12)$$

in which

$$h_e(\psi_i, \theta_i) \equiv V_{OC}/E^i = j\beta NA \cos \psi_i \sin \theta_i \quad (5-13)$$

The voltage across an arbitrary load impedance Z_L connected to the terminals of the loop with input impedance Z is determined from the Thévenin equivalent circuit in Figure 5-6:

$$V_L = V_O Z_L/(Z + Z_L) \quad (5-14)$$

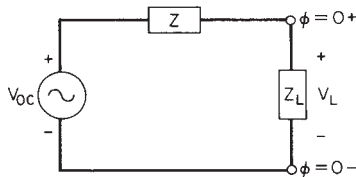


FIGURE 5-6 Thévenin equivalent circuit for the receiving loop

Ferrite-Loaded Receiving Loop

The open-circuit voltage at the terminals of the electrically small receiving loop can be increased by filling the loop with a core of permeable material, usually a ferrite. The effect of the core is to increase the magnetic flux through the area of the loop, as illustrated in Figure 5-7 for a solenoidal coil with a cylindrical core placed in a uniform axial magnetic field.

The ferrite material is characterized by a complex relative initial permeability $\mu_r = \mu/\mu_0 = \mu'_r - j\mu''_r$ and a relative permittivity $\epsilon_r = \epsilon/\epsilon_0$.^{*} The material is usually selected to have a loss tangent $p_m = \mu''_r/\mu'_r$ that is small at the frequency of operation. Consequently, μ''_r is ignored in the analysis except when the power dissipated in the core is being calculated. The dimensions of the core are also assumed to be small compared with the wavelength in the ferrite $\lambda_m \approx \lambda/\sqrt{\epsilon_r \mu'_r}$ to prevent internal resonances within the core.⁷

The open-circuit voltage for a single-turn loop at the middle of a ferrite cylinder of length ℓ_r and radius b is increased by the factor μ_{rod} over the value for the same loop in free space:

$$V_{\text{OC}} = j\omega\mu_{\text{rod}}AB_z^i \quad (5-15)$$

Here the radius of the loop conductor a is ignored, and the mean radius of the loop and the core are assumed to be the same value b . The graph in Figure 5-8 shows the apparent permeability μ_{rod} as a function of the length-to-diameter ratio for the rod $\ell_r/2b$ with the relative initial permeability of the ferrite μ'_r as a parameter.⁸ Similar graphs for the apparent permeability of solid and hollow spheroidal cores are in the literature.⁹

For a single-layer solenoidal coil of length ℓ_c centered on the rod, an averaging factor F_V must be included in the open-circuit voltage to account for the decrease in the flux along the length of the coil from the maximum at the middle:

$$V_{\text{OC}} = j\omega\mu_{\text{rod}}F_VNAB_z^i \quad (5-16)$$

The empirical factor F_V , determined from an average of experimental results, is shown in Figure 5-9 as a function of the ratio ℓ_c/ℓ_r (length of the coil to length of the rod).^{10,11} For a long rod of moderate permeability ($\ell_r/2b \gg 1$, $\mu_{\text{rod}} \approx \mu'_r$) covered by a coil of equal length ($\ell_c/\ell_r = 1$), the open-circuit voltage is increased by approximately the factor 0.8 μ'_r over the open-circuit voltage for the same coil without the core.

The equivalent circuit for the impedance of the ferrite-loaded solenoidal coil is that shown in Figure 5-3 with an additional series resistor R^m included to account for the power dissipated in the core. The elements in the circuit are the radiation resistance:

$$R^r = \frac{\zeta}{6\pi} \beta^4 (\mu_{\text{rod}} F_V N A)^2 \quad (5-17)$$

^{*}The initial permeability is the derivative dB/dH in the limit as H is reduced to zero. Dielectric loss in the ferrite is ignored here, and the permittivity is assumed to be real.

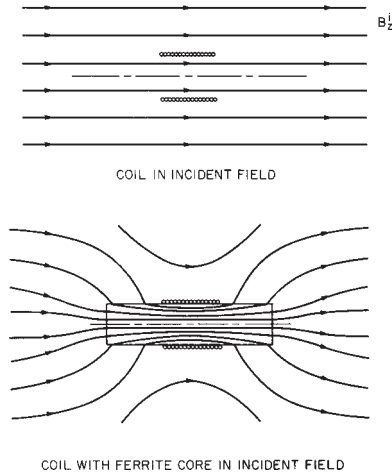


FIGURE 5-7 Effect of a cylindrical ferrite core on the magnetic flux through a solenoidal coil

the resistance due to core loss:

$$R^m = \omega(\mu_{\text{rod}}/\mu_r')^2 \mu_r'' \mu_0 F_R N^2 A / \ell_c \quad (5-18)$$

and the external inductance of the loaded solenoidal coil:

$$L^e = \mu_{\text{rod}} F_L \mu_0 N^2 A / \ell_c \quad (5-19)$$

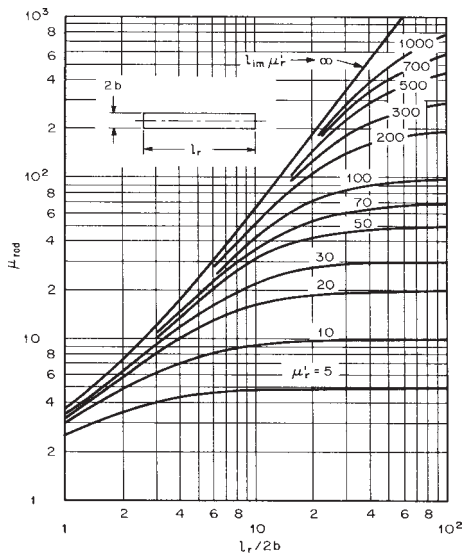


FIGURE 5-8 The apparent permeability μ_{rod} at the middle of a cylindrical rod as a function of the length-to-diameter ratio $\ell_r/2b$ with the initial permeability μ_r' as a parameter

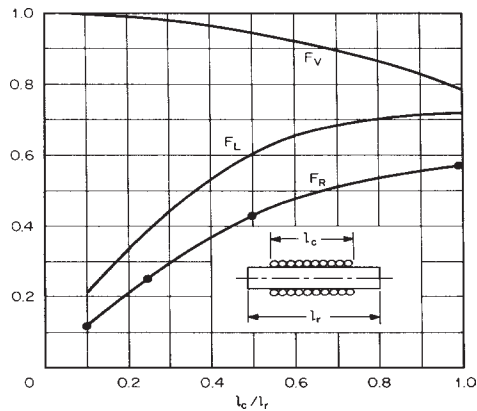


FIGURE 5-9 The factors F_V , F_L , and F_R as functions of the ratio ℓ_c/ℓ_r (length of the coil to length of the rod). These factors were determined from averages of experimental data.

The internal impedance of the conductor Z^i is assumed to be the same as that for the unloaded loop. The empirical factors F_R and F_L in Eqs. 5-18 and 5-19, like F_V , were determined from an average of experimental results and are also shown as a function of the ratio ℓ_c/ℓ_r in Figure 5-9.^{10,11} It should be emphasized that the graphs for the three factors F_V , F_R , and F_L represent typical measured values and show only the dependence on the ratio ℓ_c/ℓ_r ; some dependence on the other parameters describing the coil and the rod is to be expected.

Equations 5-16 through 5-19 provide a complete description of the electrically small ferrite-loaded receiving loop (single-layer solenoidal coil with a cylindrical core). Other parameters of interest, such as the Q of the antenna, can be determined from these results. The permeability of a specific ferrite can be obtained from the manufacturer or from the extensive tables and charts in *Soft Ferrites: Properties and Applications* by E. C. Snelling.¹¹ The many parameters that are to be chosen for the ferrite-loaded loop, such as μ_r , ℓ_r , ℓ_c , N , etc., offer a great deal of flexibility in its design. Several discussions in the literature determine these parameters to optimize the performance for a particular application.¹²

The electromagnetic field of the ferrite-loaded transmitting loop is given by Eqs. 5-1 to 5-3 with the moment $m = \mu_{\text{rod}} F_V I_0 N A$. The ferrite-loaded loop, however, is seldom used as a transmitting antenna because of the problems associated with the nonlinearity and the dissipation in the ferrite at high magnetic field strengths.¹³

Matching Networks

A matching network is generally used to improve the power transfer to and from the electrically small loop antenna. At the frequency ω_0 , the impedance of the loop can be transformed to the resistance R_0 by using various networks, two of which are shown in Figure 5-10.¹⁴ In both arrangements, the capacitor C_1 is used to bring the loop close to resonance at ω_0 . Then the additional elements (the capacitor C_2 in Figure 5-10a, or the spacing s_2 and length of the lines l_2 in Figure 5-10b) are adjusted to make the impedance R_0 . The matches obtained with these simple networks generally have very narrow bandwidth.

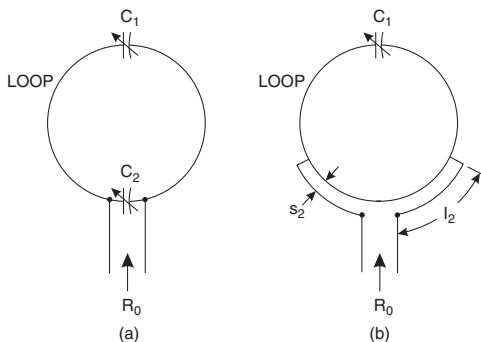


FIGURE 5-10 Examples of matching networks used with electrically small loop antennas. Matching to the resistance R_0 with (a) two capacitors and with (b) shunt feed and one capacitor.

The capacitors for the circuit in Figure 5-10a are

$$C_1 = \frac{1}{\omega_0 (\omega_0 L - R \sqrt{R_0/R - 1})} \approx \frac{1}{\omega_0 (\omega_0 L - \sqrt{RR_0})} \quad (5-20)$$

$$C_2 = \frac{1}{\omega_0 R_0 \sqrt{R_0/R - 1}} \approx \frac{1}{\omega_0 \sqrt{RR_0}} \quad (5-21)$$

in which the approximations on the right apply for the practical case $R_0/R \gg 1$. Figure 5-10b is for a symmetrical shunt-driven loop. Sometimes an asymmetrical configuration is used, in which the length of one of the two lines is set to zero.

5.3 ELECTRICALLY LARGE LOOPS

As the electrical size of the loop antenna is increased, the current distribution in the loop departs from the simple uniform distribution of the electrically small loop. For single-turn loops, this departure has a significant effect on performance when the circumference is greater than about 0.1λ . For example, the radiation resistance of an electrically small circular loop with a uniform current, as predicted by Eq. 5-5, is about 86 percent of the actual resistance when $\beta b = 2\pi b/\lambda = 0.1$ and only about 26 percent of the actual resistance when $\beta b = 0.3$.

Of the possible shapes for an electrically large loop antenna, the single-turn thin-wire circular loop has received the most attention, both theoretical and experimental. The popularity of the circular loop is due in part to its straightforward analysis by expansion of the current in the loop as a Fourier series:

$$I(\phi) = I_0 + 2 \sum_{n=1}^m I_n \cos n\phi \quad (5-22)$$

in which the angle ϕ is defined in Figure 5-1.¹⁵ Measurements on electrically large loops with other shapes, such as the square loop, show that their electrical performance is qualitatively similar to that of the circular loop; therefore, only the circular loop will be discussed here.¹⁶

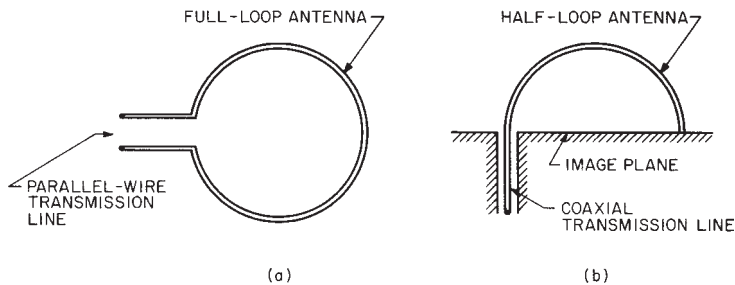


FIGURE 5-11 Methods of driving the circular-loop antenna: (a) Full-loop antenna driven from parallel-wire transmission line; (b) Half-loop antenna driven from coaxial transmission line

Circular-Loop Antenna

The theoretical model for the circular-loop antenna generally assumes a point-source generator of voltage V at the position $\phi = 0$, making the input impedance of the loop $Z = R + jX = VI(\phi = 0)$. In practical applications, the full-loop antenna is usually driven from a balanced source, such as a parallel-wire transmission line, and the half-loop antenna, the analog of the electric monopole, is driven from a coaxial line, as in Figure 5-11. The point-source generator of the theoretical model contains no details of the geometry of the feed point, and it is not strictly equivalent to either of these methods of excitation. However, theoretical current distributions, input impedances, and field patterns computed with the point-source generator and 20 terms in the Fourier series (Eq. 5-22) are generally in good agreement with measured values.* Thus, the theory serves as a useful design tool.

For the half-loop antenna (Figure 5-11b), an accurate analysis, based on the Fourier series, is available that includes the details of the feeding coaxial line.¹⁷ Results from this analysis are in excellent agreement with measurements and show such things as the change in the input impedance of the half-loop antenna with a change in the characteristic impedance of the feeding coaxial line.

In Figures 5-12 and 5-13, the input impedance of a loop constructed from a perfect conductor is shown as a function of the electrical size of the loop $\beta b = 2\pi b/\lambda$ (circumference/wavelength) for various values of the radius of the conductor, indicated by the *thickness parameter* $\Omega = 2 \ln(2\pi b/a)$. These impedances are for full-loop antennas; for half-loop antennas with the same radius and conductor size, impedances are approximately one-half of these values. The reactance X is seen to be zero at points near $\beta b = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$ (antiresonant points) and $\beta b = 1, 2, 3, \dots$ (resonant points). The resistance obtains relative maxima near the points of antiresonance and relative minima near the points of resonance. Impedances computed from Eqs. 5-5 and 5-7, which apply to electrically small loops, are also shown in Figures 5-12 and 5-13; the inaccuracy of these formulas with increasing βb is evident.

When the electrical size of the loop is near that for resonance ($\beta b = 1, 2, 3, \dots$), the dominant term in the Fourier series for the current (Eq. 5-22) is the one with $n = \text{integer } (\beta b)$. For example, near the first resonance $\beta b \approx 1$, the current in the loop is approximately $I(\phi) = 2I_1 \cos \phi$, and the loop is commonly referred to as a resonant loop. The resonant loop ($\beta b \approx 1$) is the most frequently used electrically large loop. It has a reasonable input resistance, $R \approx 100 \Omega$, for matching to a transmission line, particularly when compared with the resistance of the antiresonant loop ($\beta b \approx 0.5$), which may be larger than $10 \text{ k}\Omega$.

*The theoretical results in Figures 5-12, 5-13, 5-15 to 5-19, and 5-21 were computed by the author by using 20 terms in this series.

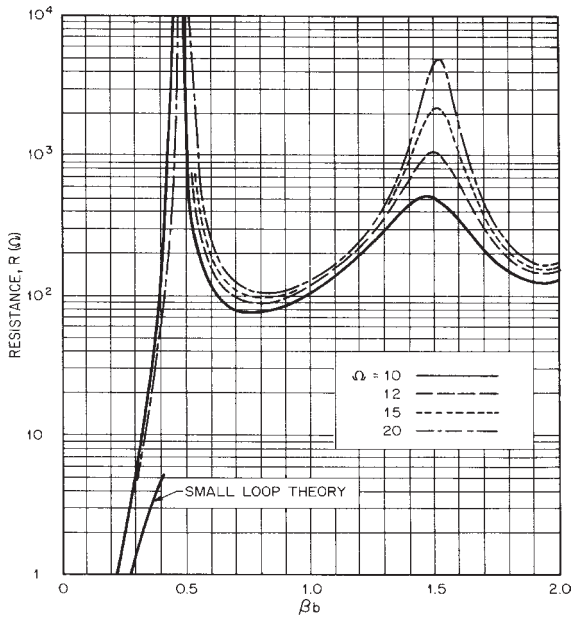


FIGURE 5-12 Input resistance of circular-loop antenna versus electrical size (circumference/wavelength)

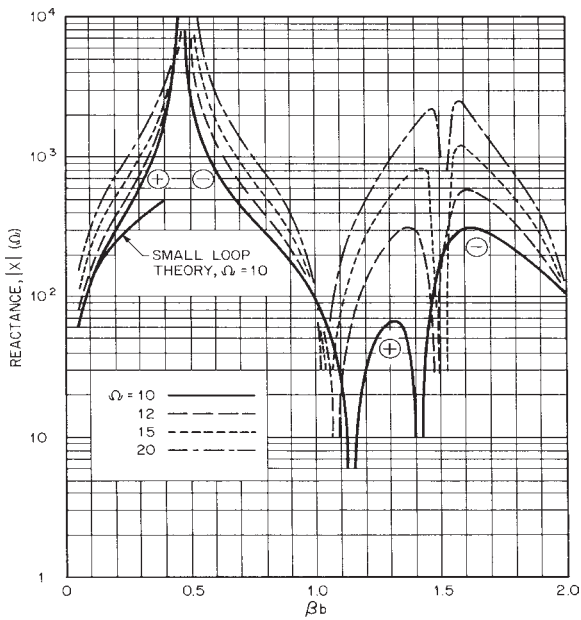


FIGURE 5-13 Input reactance of circular-loop antenna versus electrical size (circumference/wavelength)

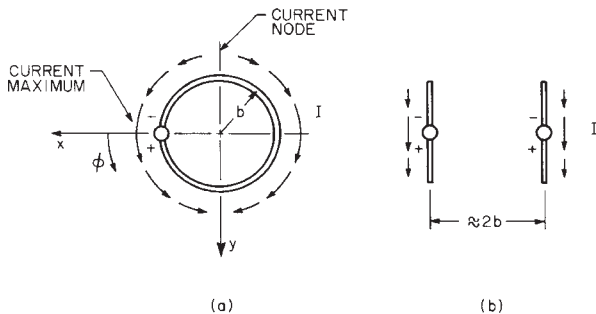


FIGURE 5-14 Schematic of current distribution in (a) resonant loop and in (b) the approximately equivalent pair of dipoles

Resonant Circular Loop

The current in the resonant loop has maxima at the generator, $\phi = 0$, and at the diametrically opposite point, $\phi = \pi$, with nodes at $\phi = \pi/2$ and $3\pi/2$. On examination of Figure 5-14, the current is seen to be roughly equivalent to that in a pair of parallel dipole antennas driven in phase and with a spacing approximately equal to the diameter of the loop.

The far-zone field patterns for the resonant loop shown in Figure 5-15a-c are also similar to those for the pair of dipoles; they have little resemblance to the figure-eight pattern of the electrically small loop, Figure 5-2. There are two components to the electric field, E_θ and E_ϕ . E_θ is zero in the horizontal plane $\theta = \pi/2$ and in the vertical plane $\phi = 0, \pi$, while E_ϕ is small in the vertical plane $\phi = \pi/2, 3\pi/2$. The amplitude patterns are symmetrical about

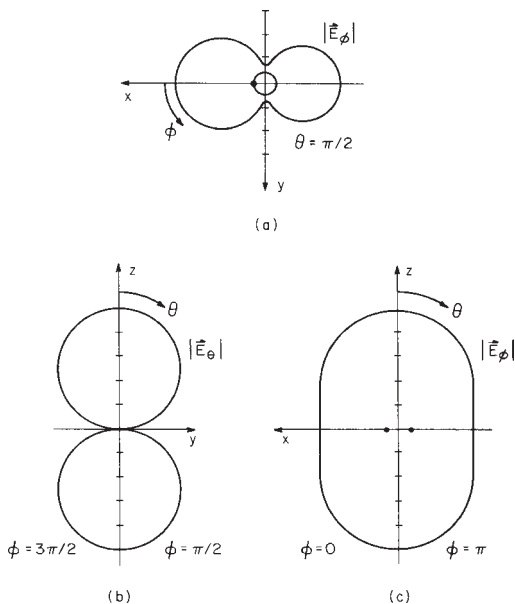


FIGURE 5-15 Far-zone electric field for loop with $\beta b = 1.0$, $\Omega = 10$: (a) Horizontal-plane field pattern $|E_\phi|$, $\theta = \pi/2$; (b) Vertical-plane field pattern $|E_\theta|$, $\phi = \pi/2, 3\pi/2$; (c) Vertical-plane field pattern $|E_\phi|$, $\phi = 0, \pi$

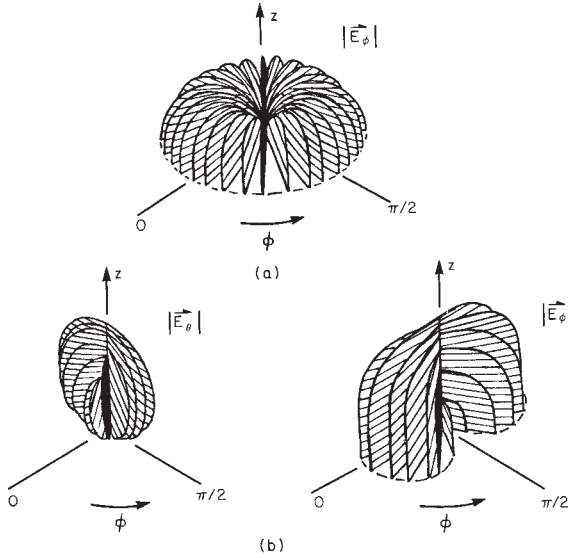


FIGURE 5-16 Far-zone electric field patterns in upper hemisphere: (a) Electrically small loop, $\beta b = 1$; (b) Resonant loop, $\beta b = 1.0$

the plane $\theta = \pi/2$ and $\phi = 0, \pi$ owing to the geometrical symmetry of the loop, and they are nearly symmetrical about the plane $\phi = \pi/2, 3\pi/2$, because of the dominance of the term $2I_1 \cos \phi$ in the current distribution. At the maxima ($\theta = 0, \pi$) of the bidirectional pattern, the electric field is linearly polarized in the direction \hat{y} .

To help with the visualization of the electric field, three-dimensional amplitude patterns for the electrically small loop and the resonant loop are presented in Figure 5-16. Each drawing is a series of patterns on planes of constant angle ϕ ; only the patterns in the upper hemisphere ($0 \leq \theta \leq \pi/2$) are shown, since those in the lower hemisphere are identical.

The directivity of the circular loop in the direction $\theta = 0$ or π is shown as a function of the electrical size βb in Figure 5-17; it is about 3.4 dB for $\beta b = 1.0$ and has a maximum of about 4.5 dB for $\beta b = 1.4$. The directivity is fairly independent of the parameter Ω for $\beta b \leq 1.4$.

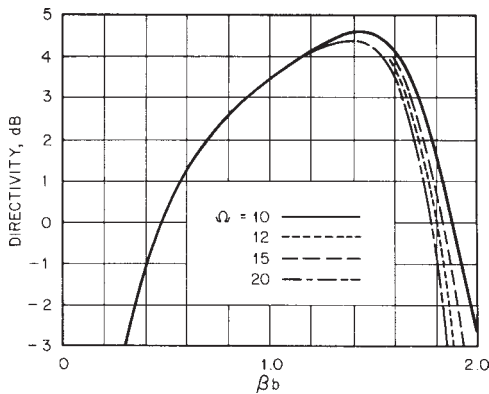


FIGURE 5-17 Directivity of circular-loop antenna for $\theta = 0, \pi$ versus electrical size (circumference/wavelength)

The resonant loop antenna is attractive for practical applications because of its moderate input resistance and symmetrical field pattern with reasonable directivity. The bidirectional nature of its pattern, however, is usually not desired, and a reflector or an array of loops is used to make the pattern unidirectional.

For most antenna problems, the far-zone field described previously is of interest; however, for some applications the field close to the loop may be required. Expressions for the near field of the circular-loop antenna with uniform current or the Fourier series representation in Eq. 5-22 have been obtained using various methods.¹⁸

Circular Loop with Planar Reflector

The pattern of the resonant loop is made unidirectional and the directivity in the direction $\theta = 0$ is increased by placing the loop over a planar reflector. The theoretical results for an infinite perfectly conducting reflector (Figure 5-18) show that the directivity is greater than 9 dB for spacings between the loop and the reflector in the range $0.05 \leq d/\lambda \leq 0.2$.¹⁹ Over this same range of spacings, the input impedance $Z = R + jX$ (Figure 5-19) has values that are easily matched: the resistance is reasonable ($R \leq 135 \Omega$), and the reactance is small ($|X| \leq 20 \Omega$).

The theoretical results for an infinite reflector are in good agreement with measured data for finite square reflectors of side length s . The directivities measured by Adachi and Mushiake²⁰ (Figure 5-18) for a reflector with $s/\lambda = 1.2$ and $d/\lambda \leq 0.26$ are slightly higher than those for an infinite plane, while the input impedances measured by Rojarayanont and Sekiguchi²¹ (Figure 5-19) show variations with reflector size, $0.48 \leq s/\lambda \leq 0.95$, but general agreement with the results for an infinite plane.

Electric field patterns measured by Rojarayanont and Sekiguchi²¹ for resonant loops one-quarter wavelength, $d/\lambda = 0.25$, in front of square reflectors are shown in Figure 5-20. The shaded area in each figure shows the variation in the pattern that is a result of changing the size of the square reflector from $s/\lambda = 0.64$ to $s/\lambda = 0.95$.

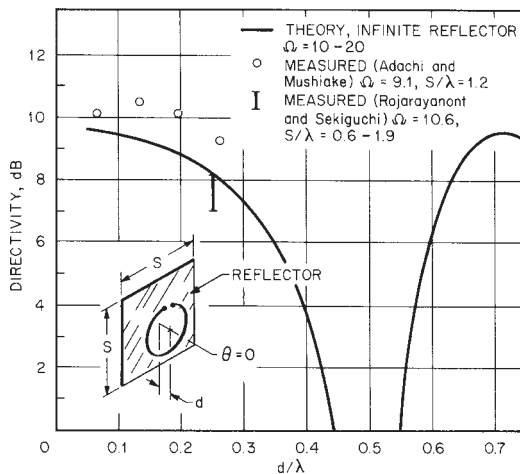


FIGURE 5-18 Directivity of circular-loop antenna, $\beta b = 1.0$, for $\theta = 0$ versus distance from reflector d/λ . Theoretical curve is for infinite planar reflector; measured points are for square reflector.

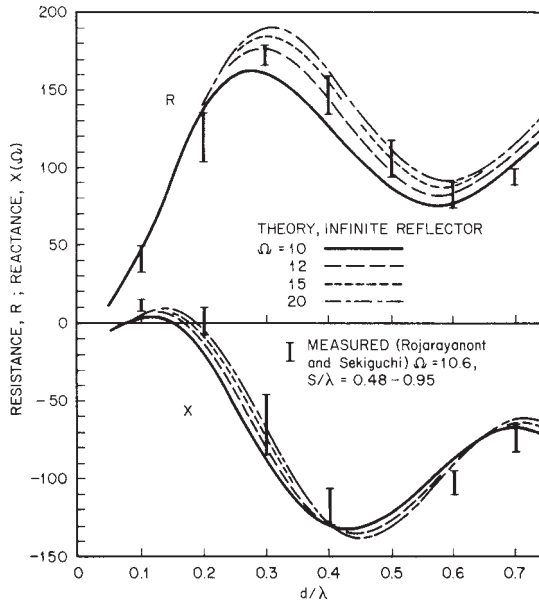


FIGURE 5-19 Input impedance of circular-loop antenna, $\beta b = 1.0$, versus distance from reflector d/λ . Theoretical curves are for infinite planar reflector; measured points are for square reflector.

Coaxial Arrays of Circular Loops

Loop antennas, like linear antennas, can be combined in an array to improve performance. The most common array of circular loops is the coaxial array in which all the loops are parallel and have their centers on a common axis; an example of a coaxial array

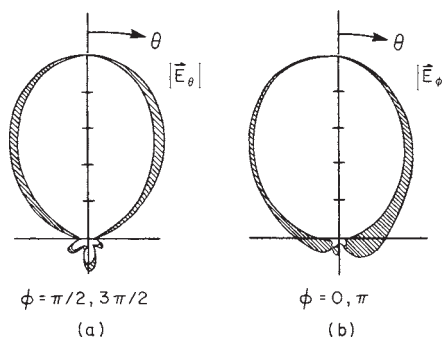


FIGURE 5-20 Measured far-zone electric field patterns for loop with $\beta b = 1.0$ over square reflector, $d/\lambda = 0.25$. Inner curve $s/\lambda = 0.95$; outer curve $s/\lambda = 0.64$. (a) Vertical-plane field pattern $|E_\theta|$, $\phi = \pi/2, 3\pi/2$. (b) Vertical-plane field pattern $|E_\phi|$, $\phi = 0, \pi$. (Measured data from Rojarayanont and Sekiguchi)

is shown later in the inset of Figure 5-22. The Fourier-series analysis for the single loop is easily extended to the coaxial array when all the driven loops are fed at a common angle, e.g., $\phi = 0$ in Figure 5-1. The current distribution in each loop is expressed as a series of trigonometric terms like that in Eq. 5-22. The simplicity of the analysis results from the orthogonality of the trigonometric terms, which makes the coupling between loops occur only for terms of the same order n . Thus, if all the driven loops in the array are near resonant size, $\beta b \approx 1$, the term $n = 1$ is the dominant one in the current distributions for all loops; i.e., the current is approximately proportional to $\cos \phi$ in all loops.

When all the elements in the loop array are driven, the same procedures that are used with arrays of linear elements can be applied to select the driving-point voltages to optimize certain parameters, such as directivity.²² The feed arrangement needed to obtain the prescribed driving-point voltages, however, is very complex for more than a few elements in the array. As a result, a simpler and more economical arrangement, an array containing only one driven element and several parasitic loops, is often used (a *parasitic loop* is a continuous wire with no terminals).

When a single, closely spaced parasite is used with a driven loop, the parasite may act as a director or as a reflector. This is illustrated in Figure 5-21, in which electric field patterns are shown for a driven loop ($\beta b_2 = 1.0$) and a parasitic loop with the spacing $d/\lambda = 0.1$. For loops of the same electrical size ($\beta b_1 = \beta b_2 = 1.0$), the maxima in the pattern at $\theta = 0, \pi$ are nearly equal. The parasitic loop that is slightly smaller than the driven loop ($\beta b_1 = 0.95$) acts as a director, producing a maximum in the pattern at $\theta = \pi$, while the parasitic loop that is slightly larger than the driven loop ($\beta b_1 = 1.05$) acts as a reflector, producing a maximum in the pattern at $\theta = 0$. This behavior is very similar to that observed for a resonant linear antenna with a closely spaced parasite.

The driven loop of electrical size $\beta b_2 = 1.2$ ($\Omega_2 = 11$, $a_1 = a_2$) with a single parasite was studied in detail by Ito et al.²³ In that study, the optimum director was determined to be a loop with $\beta b_1 \approx 0.95$ and spacing $d/\lambda \approx 0.10$. This produced a directivity of about 7 dB at $\theta = \pi$. The optimum reflector was a loop with $\beta b_1 \approx 1.08$ and a spacing $d/\lambda \approx 0.15$. This produced a directivity of about 8 dB at $\theta = 0$. Note that, for this case, the optimum director and the optimum reflector are both smaller than the driven loop.

A Yagi-Uda array of loops with a single reflector (element 1), an exciter (the driven element 2), and several directors of equal size βb and equal spacing d/λ is shown in the

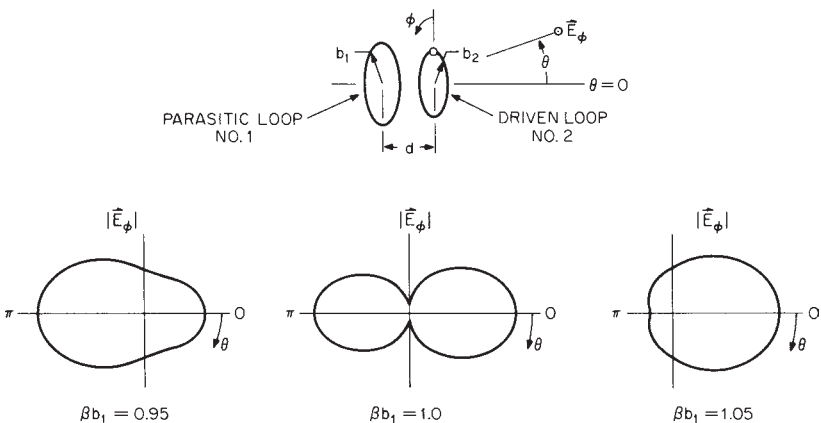


FIGURE 5-21 Far-zone electric field patterns $|\mathbf{E}_\phi|$ in plane $\phi = 0, \pi$ for driven loop with single parasite, $\beta b_2 = 1.0$, $d/\lambda = 0.1$, $\Omega_1 = \Omega_2 = 20$

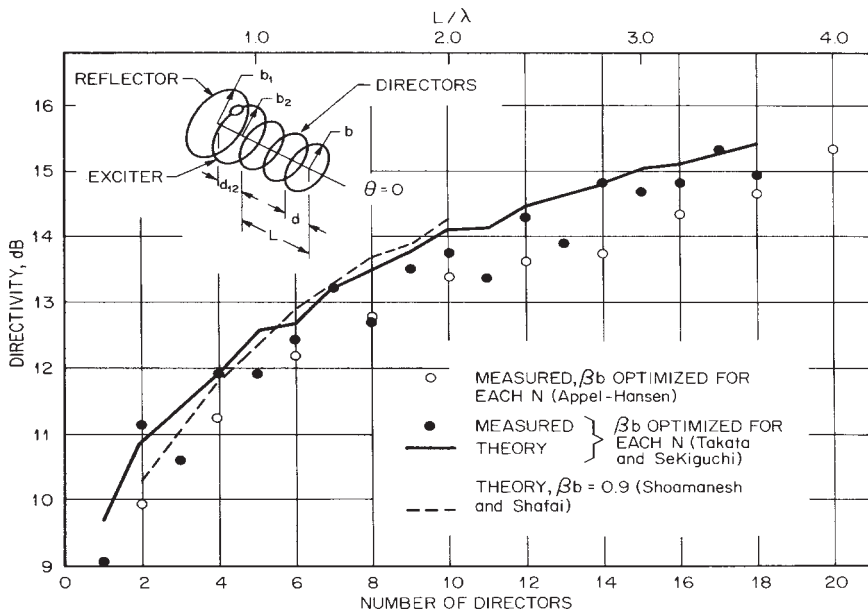


FIGURE 5-22 Directivity of Yagi-Uda array of circular-loop antennas for $\theta = 0$ versus number of directors, director spacing $d/\lambda = 0.2$

inset of Figure 5-22.* As in its counterpart with linear elements, in the Yagi-Uda array of loops the reflector-exciter combination acts as a feed for a slow wave that propagates along the array of directors.²⁴ The lowest-order propagating wave (mode) exists for directors less than about resonant size ($\beta b \leq 1.0$) with spacings less than about a half wavelength ($d/\lambda \leq 0.5$).²⁵ An array supporting this mode has an end-fire pattern with a linearly polarized electric field at the maximum, $\theta = 0$.

The procedure for designing a Yagi-Uda array of loops is the same as for an array with linear elements.²⁶ The isolated reflector-exciter combination is usually chosen to have maximum directivity in the direction $\theta = 0$. For example, the optimized two-element array described previously might be used. The number, size, and spacing of the directors are then adjusted to obtain the desired performance, such as specified end-fire directivity. The maximum end-fire directivity is determined by the electrical length of the array L/λ (L is the distance from the exciter to the last director). The larger the number of directors within the length L , the smaller the electrical size of the directors will be for maximum directivity, typically $0.8 \leq \beta b \leq 1.0$.

As an example, the directivity of a Yagi-Uda array of loops with the director spacing $d/\lambda = 0.2$ is shown as a function of the number of directors or the length of the array L/λ in Figure 5-22. Two theoretical curves and two sets of measured data are shown. All the results agree to within about 1 dB, even though they are for different reflector-exciter combinations and slightly different director sizes.[†]

*In the literature of amateur radio the Yagi-Uda array of loops, usually square loops, is referred to as a *quad antenna*.

†The parameters used by the different investigators are Appel-Hansen, $\beta b_1 = \beta b_2 = 1.10$, d_{12}/λ optimized for the isolated reflector-exciter, and βb optimized for each length L/λ ; Takata and Sekiguchi, $\beta b_1 = 1.05$, $\beta b_2 = 1.20$, $d_{12}/\lambda = 0.15$, and βb optimized for each length L/λ ; Shoamanesh and Shafai (1979), $\beta_1 b = 1.05$, $\beta b_2 = 1.10$, $d_{12}/\lambda = 0.1$, and $\beta b = 0.9$ for all lengths L/λ .

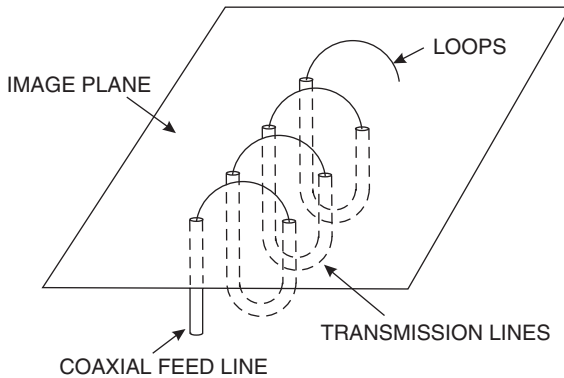


FIGURE 5-23 Coaxial array of half-loop antennas. Spacing between loops and lengths of interconnecting transmission lines can be adjusted to change field pattern of the array.

Arrays can also be formed using the half-loop antenna shown in Figure 5-11*b* as the element.²⁷ The coaxial loops can be fed individually, used as parasites, or they can be connected in series by transmission lines below the image plane and fed at a single point, as in Figure 5-23. In this arrangement, the spacing between the loops and the lengths of the transmission lines can be adjusted to change the field pattern of the array.

5.4 SHIELDED-LOOP ANTENNA

For certain applications it is desirable to position the terminals of the loop antenna precisely so as to produce geometrical symmetry for the loop and its connections about a plane perpendicular to the loop. This can often be accomplished by using the so-called shielded loop. Figure 5-24*a* is an example of a shielded receiving loop whose external surface is symmetrical about the yz plane.²⁸

With reference to Figure 5-24*a*, the thickness of the metal forming the shield is chosen to be several skin depths; this prevents any direct interaction between the currents on the internal and the external surfaces of the shield. The effective terminals of the loop antenna are at the ends of the small gap AB . The inner conductor and the shield form a coaxial transmission line of length $h + \pi b$ connecting the gap with the load impedance Z_L . Thus, the effective load impedance Z_L at the gap is Z_S transformed by the length of transmission line $h + \pi b$.

The receiving antenna in Figure 5-24*a* is easily analyzed by considering the loop, Figure 5-24*b*, and the transmission line, Figure 5-24*c*, separately. The incident field produces a current on the external surface of the shield; the current passes through the effective impedance Z_L , producing the voltage V_L , which for an electrically small loop can be determined from Eqs. 5-12 and 5-14. This voltage is transmitted over the coaxial line to become V_S at the load impedance Z_S .

Other examples of the shielded loop are shown in Figure 5-25. A balanced version of the loop in Figure 5-24*a* is in Figure 5-25*a*, and a method for feeding a loop in front of a planar reflector is in Figure 5-25*b*.

To illustrate a typical use of the shielded loop, consider the electrically small receiving loop placed in an incident linearly polarized electromagnetic plane wave with the wave

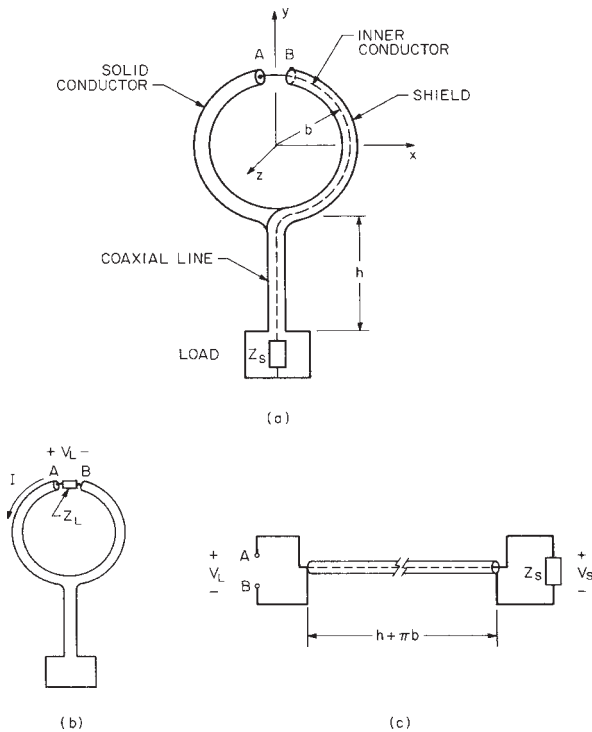


FIGURE 5-24 (a) Shielded-loop antenna (b) with equivalent antenna and (c) equivalent transmission line

vector \mathbf{k}_i , as in Figure 5-26. This is the same geometry as in Figure 5-5, except that the terminals of the loop are at the angle $\phi = \phi_L$ instead of $\phi = 0$, and $\phi_i = \pi$, $\psi_i = 0$. The loop in this example might be an antenna in a direction finder with the direction of the incident wave to be determined by placing a null of the field pattern in the direction of \mathbf{k}_i .

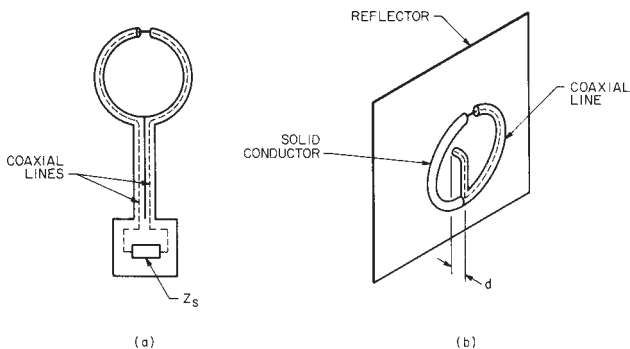


FIGURE 5-25 (a) Balanced shielded-loop antenna and (b) method of feeding loop antenna in front of planar reflector

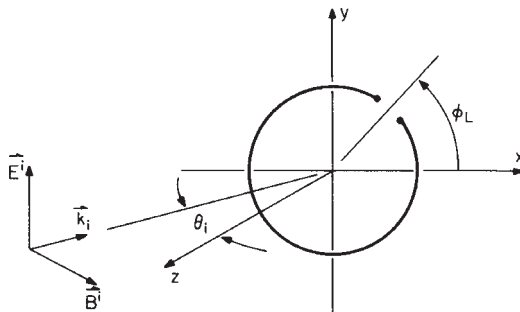


FIGURE 5-26 Receiving loop in plane-wave incident field

The voltage at the open-circuited terminals of the electrically small loop, determined from the Fourier-series analysis, is approximately*

$$V_{OC} = j\omega AB^i (\sin \theta_i - 2j\beta b \cos \phi_L) \quad (5-23)$$

For many applications, the second term in Eq. 5-23 is negligible, since $\beta b \ll 1$ for an electrically small loop. In other applications, however, this term may represent a significant contribution to the response. For example, the sensitivity of the antenna in the direction finder is decreased by this term because it fills in the nulls of the $\sin \theta_i$ field pattern (for $\beta b = 0.1$, $\phi_L = 0$, the minima in the pattern are only 14 dB below the maxima).

The second term in Eq. 5-23 can be made insignificant by reducing the electrical size of the loop βb ; however, this will also decrease the sensitivity, because the area of the loop is decreased. An alternative is to make this term zero by placing the terminals of the loop precisely at $\phi_L = \pm \pi/2$ ($\cos \phi_L = 0$). This can be accomplished by using a shielded loop as in Figure 5-24a or Figure 5-25a.

5.5 PULSE-EXCITED CIRCULAR-LOOP ANTENNA

In the treatment presented so far, the performance of the loop antenna has been described for harmonic time dependence, which is the case for most practical applications. However, there are instances when the loop antenna is used with pulse excitation, such as in some proposed ultra-wideband communications systems. With pulse excitation, the time delay between pulses can be used to identify the radiation from different points on the loop, and this provides some physical insight into how the loop radiates energy. The details for the calculations are in the literature,²⁹ and only a qualitative description for the radiation will be given here.³⁰

When the circular loop is excited by a voltage pulse, such as a Gaussian, pulses of charge travel around the loop at roughly the speed of light, c : a positive pulse travels in the clockwise direction, and a negative pulse travels in the counterclockwise direction, as shown in Figure 5-27a. When these pulses have traveled once around the loop, they return to the source, where they are partially absorbed and partially transmitted, as shown in Figure 5-27b. This process is repeated until there is no more charge left on the loop.

*Eq. 5-23 is a more accurate formula than Eq. 5-12, which is also for the open-circuited voltage of the electrically small loop. It contains an additional term not in Eq. 5-12, the second term in the parentheses, which is proportional to βb . When $\beta b \ll 1$, Eq. 5-23 reduces to Eq. 5-12 with $N = 1$, and $\psi_i = 0$.

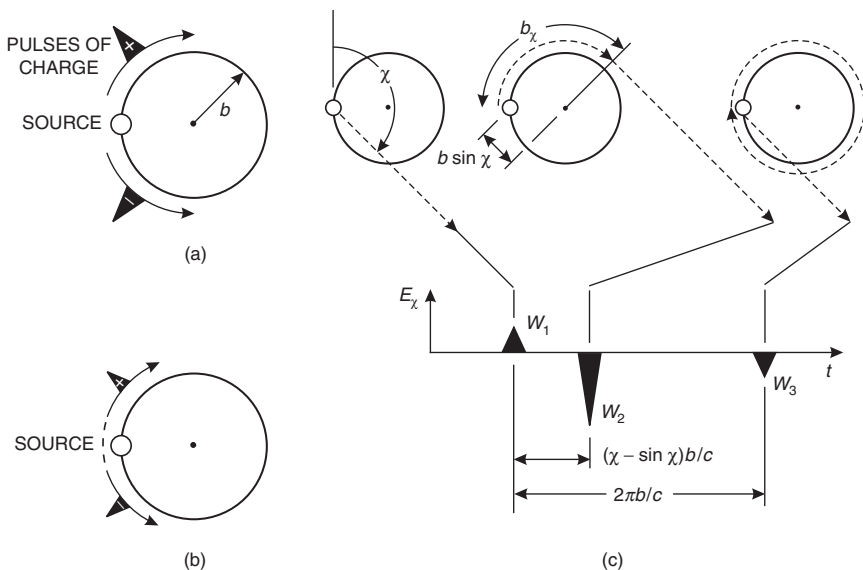


FIGURE 5-27 Schematic drawings for pulse-excited circular-loop antenna. (a) Pulses of charge leaving source and traveling around loop. (b) Pulses of charge after completing one trip around loop and passing through source. (c) Time history of far-zone electric field radiated by loop. Results are shown in the plane of the loop for the radiation due to the positive pulse of charge making one trip around the loop.

The far-zone electric field produced by one of these pulses of charge, the positive one traveling in the clockwise direction once around the loop, is described schematically in Figure 5-27c. For this example, the width of the pulse is small compared with the time for light to travel around the perimeter of the loop ($2\pi b/c$), and the time history of the field is shown for one angle of observation in the plane of the loop, $\chi = 3\pi/4 = 135^\circ$.

Three wavefronts can be distinguished in Figure 5-27c.³⁰ The wavefronts W_1 and W_3 are produced when the pulse of charge leaves the source and when it is absorbed/transmitted at the source, respectively. The wavefront W_2 is continuously produced as the pulse propagates around the loop. Wavefront W_2 appears to originate at a point on the loop where the tangent line to the loop points in the direction of the observer. As indicated in the figure, this can be ascertained from the time of arrival of W_2 relative to W_1 . The radiation from the negative pulse of charge in Figure 5-27a and from all of the subsequent pulses, Figure 5-27b, can be obtained in a similar manner.

A simple analogy can be constructed between the radiation from this loop and the radiation from a moving point charge.³⁰ The radiation arising at the source, wavefronts W_1 and W_3 , is analogous to the radiation that occurs when a point charge is accelerated in a direction parallel to its velocity (*bremsstrahlung*), and the radiation that occurs along the loop, wavefront W_2 , is analogous to the radiation that occurs when a point charge is accelerated in a direction normal to its velocity (*synchrotron radiation*).

When the loop is excited with a time-harmonic signal, a cosinusoid, each half of the cosinusoid can be considered to be a pulse. The radiation is then that due to a train of positive and negative pulses of charge—the positive pulses being separated in time by the period of the cosinusoid, $T = 2\pi/\omega = \lambda/c$. Notice that the charge on the loop from consecutive positive pulses will add when the travel time around the loop is T . This occurs for $\beta b = 2\pi b/\lambda = 1$, which is the condition for resonance stated earlier in connection with the input impedance of the loop.

In practical applications, it may be desirable to eliminate or reduce the amplitude of some of the pulses radiated by the loop. This can be accomplished by incorporating discrete or continuous resistive loading along the wire of the loop.^{30,31}

5.6 ADDITIONAL TOPICS

The brevity of this review requires omission of many interesting topics concerning loop antennas. In recent years there has been considerable study of loop antennas in close proximity to or embedded in material media such as the ocean, the earth, and biological tissue. The electrical characteristics of loops in these instances can be quite different from those of loops in unbounded free space, as described in this review. The major applications of this work are in the areas of subsurface communication and detection (geophysical prospecting) and the study of the interaction of electromagnetic radiation with biological systems.

The loop antenna near a planar interface separating two semi-infinite material regions, such as the air and the earth, has been investigated extensively. When the loop is electrically small, it can be approximated by an elementary magnetic dipole, and the electromagnetic field away from the loop can be determined from the classical analysis of Sommerfeld.³² If the field near the electrically small loop is required, the approximation by a magnetic dipole may no longer be adequate, and a loop with a finite radius and a uniform current must be considered.³³ For the electrically large loop near a planar interface, an analysis that allows a nonuniform current in the loop, such as the Fourier-series analysis for the circular loop,³⁴ must be used.

The performance of a loop embedded in a material can be altered significantly by placing the loop in a dielectric cavity, such as a sphere, to form an insulated loop. The electrical size and shape of the insulating cavity and the location of the loop in the cavity can be used to control the electromagnetic field and input impedance of the antenna.³⁵

Models have been developed to analyze a loop antenna placed near the human body, and the results have been used for two different purposes.³⁶ In one case the concern is the possible hazard due to the electromagnetic energy deposited in the body. In the other case the concern is the effect the body has on the performance (impedance and pattern) of a loop antenna contained in a personal communications device, such as a pager.

REFERENCES

1. H. Hertz, *Electric Waves* (London: Macmillan and Co., Ltd., 1893).
2. R. G. Medhurst, "H.F. Resistance and Self-Capacitance of Single-Layer Solenoids," *Wireless Eng.*, vol. 24 (March 1947): 80. The measurements of Medhurst show that the self-resonance of a solenoid with $\ell/2b \geq 3$ occurs at a wavelength at which $N\beta b \geq 0.4$. The current distribution in the solenoid is assumed to be uniform well below self-resonance, i.e., $N\beta b \leq 0.1$.
3. The electrically small loop antenna in free space is discussed in many textbooks; see, for example, G. S. Smith, *An Introduction to Classical Electromagnetic Radiation* (Cambridge, UK: Cambridge University Press, 1997): 477–485; R. W. P. King, *Fundamental Electromagnetic Theory* (New York: Dover Publications, Inc., 1963): 441–457; and S. A. Schelkunoff and H. T. Friis, *Antennas: Theory and Practice* (New York: John Wiley & Sons, Inc., 1952): 319–324.
4. Formulas and graphs for the internal impedance per unit length of round conductors are in S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics* (New York: John Wiley & Sons, Inc., 1965): 286–297.
5. G. S. Smith, "Radiation Efficiency of Electrically Small Multiturn Loop Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (September 1972): 656.

6. F. W. Grover, *Inductance Calculations: Working Formulas and Tables* (New York: D. Van Nostrand Company, Inc., 1946).
7. Internal resonance transverse to the axis of an infinitely long magnetic rod is discussed in L. Page, "The Magnetic Antenna," *Phys. Rev.*, vol. 69 (June 1946): 645.
8. The graph in Figure 5-8 was constructed by using the static demagnetizing factor for a cylindrical rod as presented in R. M. Bozorth and D. M. Chapin, "Demagnetizing Factors of Rods," *J. App. Phys.*, vol. 13 (May 1942): 320; also R. M. Bozorth, *Ferromagnetism* (New York: D. Van Nostrand Company, Inc., 1951): 845-849; and G. A. Burtsev, "Computing the Demagnetization Coefficient of Cylindrical Rods," *Soviet J. Nondestructive Test. (Defektoskopiya)*, vol. 5 (September-October 1971): 499.
9. The receiving loop with a spheroidal core is discussed in R. E. Burgess, "Iron-Cored Loop Receiving Aerial," *Wireless Eng.*, vol. 23 (June 1946): 172; J. R. Wait, "Receiving Properties of Wire Loop with a Spheroidal Core," *Can. J. Tech.*, vol. 31 (January 1953): 9, and "The Receiving Loop with a Hollow Prolate Spheroidal Core," *Can. J. Tech.*, vol. 31 (June 1953): 132; V. H. Rumsey and W. L. Weeks, "Electrically Small, Ferrite-Loaded Loop Antennas," *IRE Conv. Rec.*, Part 1 (1956): 165; and E. J. Scott and R. H. DuHamel, "Effective Permeability of Spheroidal Shells," Tech. Rep. 9, Antenna Lab., University of Illinois, Urbana (1956).
10. The customary procedure for analyzing the solenoidal coil with a cylindrical ferrite core is in H. van Suchtelen, "Ferrocube Aerial Rods," *Electron. Appl. Bull.*, vol. 13 (June 1952): 88. Extensions of this procedure and additional measured data are in J. S. Belrose, "Ferromagnetic Loop Aerials," *Wireless Eng.*, vol. 32 (February 1955): 41; and J. Dupuis, "Cadres utilisant des ferrites," *L'onde electrique*, vol. 35 (March-April 1955): 379.
11. E. C. Snelling, *Soft Ferrites: Properties and Applications* (Cleveland: CRC Press, 1969): 182-192, 327-336.
12. There are many journal articles in addition to those in Refs. 9 through 11 that discuss the design and optimization of ferrite-loaded loop antennas for broadcast receivers; an incomplete list follows: H. Blok and J. J. Rietveld, "Inductive Aerials for Modern Broadcast Receivers," *Philips Tech. Rev.*, vol. 16 (January 1955): 181; W. J. Polydoroff, *High-Frequency Magnetic Materials, Their Characteristics and Principal Applications* (New York: John Wiley & Sons, Inc., 1960): 163-179; E. J. Maanders and H. van der Vleuten, "Ferrite Aerials for Transistor Receivers," *Philips Matronics Tech. Info. Bull.* (February 1961): 354; H. J. Laurent and C. A. B. Carvalho, "Ferrite Antennas for A.M. Broadcast Receivers," *IRE Trans. Broadcast Telev. Receivers*, vol. BTR-8 (July 1962): 50; G. Schiefer, "A Small Ferrocube Aerial for VHF Reception," *Philips Tech. Rev.*, vol. 24 (1962-1963): 332; I. D. Stuart, "Practical Considerations in the Design of Ferrite Cored Aerials for Broadcast Receivers," *IEEE Proc. (Australia)*, vol. 27 (December 1966): 329; R. C. Pettengill, H. T. Garland, and J. P. Meindl, "Receiving Antenna Design for Miniature Receivers," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (July 1977): 528.
13. The ferrite-loaded transmitting loop is discussed in R. DeVore and P. Bohley, "The Electrically Small Magnetically Loaded Multiturn Loop Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (July 1977): 496.
14. K. H. Patterson, "Down-to-Earth Army Antenna," *Electronics* (August 1967): 111; and R. W. P. King, "The Shunt-Driven Circular Loop Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-19 (September 1971): 692.
15. The Fourier-series analysis for the circular loop has a long history dating back to the work of H. C. Pocklington in 1897 on the closed loop. Recent treatments and additional references are in R. W. P. King and G. S. Smith, *Antennas in Matter: Fundamentals, Theory, and Applications* (Cambridge: The M.I.T. Press, 1981): 527-605; and R. W. P. King, "The Loop Antenna for Transmission and Reception," in R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, part I (New York: McGraw-Hill, 1969): 458-482. Tables of input admittance are in R. W. P. King, *Tables of Antenna Characteristics* (New York: Plenum Press, 1971): 151-160. The approach used by Japanese authors is described in N. Inagaki, T. Sekiguchi, and S. Ito, "A Theory of a Loop Antenna," *Electron. Commun. Japan*, vol. 53-B (March 1970): 62.
16. P. A. Kennedy, "Loop Antenna Measurements," *IRE Trans. Antennas Propagat.*, vol. AP-4 (October 1956): 610.
17. G. Zhou and G. S. Smith, "An Accurate Theoretical Model for the Thin-Wire Circular Half-Loop Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-39 (August 1991): 1167.

18. The near field of the circular loop with uniform current is discussed in P. L. Overfelt, "Near Fields of the Constant Current Thin Circular Loop Antenna of Arbitrary Radius," *IEEE Trans. Antennas Propagat.*, vol. AP-44 (February 1996): 166, and the near field of the circular loop with a Fourier series for the current is treated by various methods in D. H. Werner, "An Exact Integration Procedure for the Vector Potentials of Thin Circular Loop Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-44 (February 1996): 157 and (August 1996): 1199; L. W. Lin, M. S. Leong, S. Kooi, and T. S. Yeo, "Exact Solutions of Electromagnetic Fields in Both Near and Far Zones Radiated by Thin Circular-Loop Antennas: A General Representation," *IEEE Trans. Antennas Propagat.*, vol. AP-45 (December 1997): 1741, vol. AP-49 (January 2001): 109, and (May 2001): 847; D. H. Werner, "Comments on Exact Solutions of Electromagnetic Fields in Both Near and Far Zones Radiated by Thin Circular-Loop Antennas: A General Representation," *IEEE Trans. Antennas Propagat.*, vol. AP-49 (January 2001): 109; M. R. Abdul-Gaffoor, H. K. Smith, A. A. Kishk, and A. W. Glisson, "Comments on Exact Solutions of Electromagnetic Fields in Both Near and Far Zones Radiated by Thin Circular-Loop Antennas: A General Representation," *IEEE Trans. Antennas Propagat.*, vol. AP-49 (May 2001): 845; and J. T. Conway, "New Exact Solution Procedure for the Near Fields of the General Thin Circular Loop Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-53 (January 2005): 509.
19. The properties of a loop over an infinite image plane are obtained by using the theory of images and the analysis for an array of two loops; see, for example, K. Iizuka, R. W. P. King, and C. W. Harrison, Jr., "Self- and Mutual Admittances of Two Identical Circular Loop Antennas in a Conducting Medium and in Air," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (July 1966): 440.
20. S. Adachi and Y. Mushiaki, "Directive Loop Antennas," *Res. Inst. Sci. Rep.*, ser. B, vol. 9, no. 2, (Sendai, Japan: Tōhoku University, 1957): 105–112.
21. B. Rojarrayanont and T. Sekiguchi, "One-Element Loop Antenna with Finite Reflector," *Electron. Commun. Japan*, vol. 59-B (May 1976): 68.
22. The coaxial array of driven loops is discussed in M. Kosugi, N. Inagaki, and T. Sekiguchi, "Design of an Array of Circular-Loop Antennas with Optimum Directivity," *Electron. Commun. Japan*, vol. 54-B (May 1971): 67; and S. Ito, M. Kosugi, N. Inagaki, and T. Sekiguchi, "Theory of a Multi-Element Loop Antenna," *Electron. Commun. Japan*, vol. 54-B (June 1971): 95.
23. S. Ito, N. Inagaki, and T. Sekiguchi, "Investigation of the Array of Circular-Loop Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-19 (July 1971): 469.
24. H. W. Ehrenspeck and H. Poehler, "A New Method for Obtaining Maximum Gain from Yagi Antennas," *IRE Trans. Antennas Propagat.*, vol. 7 (October 1959): 379.
25. M. Yamazawa, N. Inagaki, and T. Sekiguchi, "Excitation of Surface Wave on Circular-Loop Array," *IEEE Trans. Antennas Propagat.*, vol. AP-19 (May 1971): 433.
26. The design of Yagi-Uda arrays of loops is discussed in J. E. Lindsay, Jr., "A Parasitic End-Fire Array of Circular Loop Elements," *IEEE Trans. Antennas Propagat.*, vol. AP-15 (September 1967): 697; J. Appel-Hansen, "The Loop Antenna with Director Arrays of Loops and Rods," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (July 1972): 516; L. C. Shen and G. W. Raffoul, "Optimum Design of Yagi Array of Loops," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (November 1974): 829; N. Takata and T. Sekiguchi, "Array Antennas Consisting of Linear and Loop Elements," *Electron. Commun. Japan*, vol. 59-B (May 1976): 61; A. Shoamanesh and L. Shafai, "Properties of Coaxial Yagi Loop Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-26 (July 1978): 547; and A. Shoamanesh and L. Shafai, "Design Data for Coaxial Yagi Array of Circular Loops," *IEEE Trans. Antennas Propagat.*, vol. AP-27 (September 1979): 711.
27. G. P. Zhou and G. S. Smith, "The Multiturn Half-Loop Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-42 (May 1994): 750.
28. The shielded loop is discussed in L. L. Libby, "Special Aspects of Balanced Shielded Loops," *IRE Proc.*, vol. 34 (September 1946): 641; and R. W. P. King, "The Loop Antenna for Transmission and Reception," in R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, part I (New York: McGraw-Hill, 1969): 478–480. The shielded half loop as a current probe is treated in R. W. P. King and G. S. Smith, *Antennas in Matter: Fundamentals, Theory and Applications* (Cambridge: The M.I.T. Press, 1981): 770–787.

29. The loop with pulse excitation is analyzed in A. M. Abo-Zena and R. E. Beam, "Transient Radiation Field of a Circular Loop Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (May 1972): 380; G. Franceschetti and C. H. Papas, "Pulsed Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (September 1974): 651; K. J. Langenberg, "Pulsed Loop Antennas," *Appl. Phys.*, vol. 10 (1976): 309; and K. J. Langenberg, "Transient Fields of Small Loop Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-24 (March 1976): 236.
30. For more of the details for the discussion presented here, see G. S. Smith, *An Introduction to Classical Electromagnetic Radiation* (Cambridge, UK: Cambridge University Press, 1997): 567–575, and G. S. Smith, "Teaching Antenna Radiation from a Time-Domain Perspective," *Am. J. Phys.*, vol. 69 (March 2001): 288.
31. K. P. Esselle and S. S. Stuchly, "Resistively Loaded Loop as a Pulse-Receiving Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-38 (July 1990): 1123, and H. F. Li, Z. N. Chen, and L. W. Li, "Characterization of Resistive-Loaded Wire Loop in UWB (Impulse) Radio," *Microwave Opt. Technol. Lett.*, vol. 43 (October 20, 2004): 151.
32. The analysis of elementary vertical and horizontal magnetic dipoles near a planar interface is discussed in A. Sommerfeld, *Partial Differential Equations in Physics* (New York: Academic Press, Inc., 1949): 237–279; A. Baños, Jr., *Dipole Radiation in the Presence of a Conducting Half-Space* (New York: Pergamon Press, 1966); and J. R. Wait, *Electromagnetic Waves in Stratified Media* (New York: Pergamon Press, 1970).
33. J. Ryu, H. F. Morrison, and S. H. Ward, "Electromagnetic Fields about a Loop Source of Current," *Geophysics*, vol. 35 (October 1970): 862; J. R. Wait and K. P. Spies, "Subsurface Electromagnetic Fields of a Circular Loop of Current Located above Ground," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (July 1972): 520; and J. R. Wait and K. P. Spies, "Low-Frequency Impedances of a Circular Loop over a Conducting Ground," *Electron. Lett.*, vol. 9 (July 26, 1973): 346.
34. L. N. An and G. S. Smith, "The Horizontal Circular Loop Antenna near a Planar Interface," *Radio Sci.*, vol. 17 (May–June 1982): 483.
35. Bare and insulated electrically small loop antennas in dissipative media are discussed in J. R. Wait, "Electromagnetic Fields of Sources in Lossy Media," in R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, part II (New York: McGraw-Hill, 1969): 438–514 and references therein. Bare and insulated loop antennas of general size are treated in R. W. P. King and G. S. Smith, *Antennas in Matter: Fundamentals, Theory and Applications* (Cambridge: The M.I.T. Press, 1981): 527–605; and L. N. An and G. S. Smith, "The Eccentrically Insulated Circular Loop Antenna," *Radio Sci.*, vol. 15 (November–December 1980): 1067 and vol. 17 (May–June 1982): 737.
36. A. Lakhtakia, M. Iskander, C. H. Durney, and H. Massoudi, "Near-Field Absorption in Prolate Spheroidal Models of Humans Exposed to a Small Loop Antenna of Arbitrary Orientation," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29 (June 1981): 588; W. T. Chen and H. R. Chuang, "Numerical Computation of Human Interaction with Arbitrary Oriented Superquadric Loop Antennas in Personal Communications," *IEEE Trans. Antennas Propagat.*, vol. AP-46 (June 1998): 521; and W. T. Chen and H. R. Chuang, "Numerical Computation of the EM Coupling Between a Circular Loop Antenna and a Full-Scale Human-Body Model," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-46 (October 1998): 1516.

Chapter 6

Small Antennas

Steven R. Best
The MITRE Corporation

CONTENTS

6.1 THE DEFINITION OF ELECTRICALLY SMALL	6-2
6.2 THE ELECTRICALLY SMALL TRANSMIT ANTENNA	6-3
6.3 THE ELECTRICALLY SMALL DIPOLE AND LOOP ANTENNAS	6-4
6.4 POWER FACTOR, QUALITY FACTOR, AND BANDWIDTH	6-11
6.5 RADIATION EFFICIENCY AND REALIZED GAIN	6-16
6.6 THE ELECTRICALLY SMALL RECEIVE ANTENNA	6-19
6.7 THE EFFECTIVE VOLUME	6-23
6.8 THE RADIATION SHIELD (THE WHEELER CAP)	6-26
6.9 TYPICAL SMALL ANTENNAS	6-27

6.1 THE DEFINITION OF ELECTRICALLY SMALL

The definition of an electrically small antenna can be somewhat ambiguous. Often the definition of a small antenna is one having overall dimensions (including any ground plane image) less than one-quarter wavelength ($\lambda/4$), one-eighth wavelength ($\lambda/8$), or one-tenth wavelength ($\lambda/10$). In previous editions of this *Handbook* and in his work on small antennas,¹ Wheeler defined the small antenna as one whose volume occupies a small fraction of one radiansphere. The *radiansphere* is a spherical volume having a radius of $\lambda/2\pi$.² The significance of the radiansphere dimension is that it is approximately the region surrounding the small antenna primarily occupied by the stored or reactive energy of its near electric and magnetic fields.

The electrical size or volume of the small antenna is defined by the value of ka , where k is the free-space wavenumber $2\pi/\lambda$, and a is the radius of an imaginary sphere circumscribing the maximum dimension of the antenna, as illustrated in Figure 6-1. Figure 6-1 illustrates the definition of a for both a free-space dipole-like antenna and a monopole-like antenna that requires a ground plane for operation. If the finite ground plane is sufficiently large so the monopole's impedance is similar to that exhibited by the monopole on an infinite ground plane, the dimensions of the ground plane do not need to be included in the definition of a . However, if the size and shape of the finite ground plane or the location of the monopole on the ground plane are such that the monopole's impedance is remarkably different from that exhibited by the monopole on an infinite ground plane, the dimensions of the ground plane structure generally need to be included in the definition of a .

Wheeler's statement that the small antenna is one occupying a small fraction of the radiansphere has led to the commonly accepted electrical size limit for an electrically small antenna being $ka \leq 1$. Examination of Wheeler's work^{3,4} and the work of Adler et al⁵ reveals that they describe the small antenna as having a maximum dimension less than the radian length, $\lambda/2\pi$. From the definition of a in Figure 6-1 being one-half the antenna's maximum dimension, this defines a value of $ka = 0.5$ as the electrical size limit for the electrically small antenna. You could argue that this lower value of ka simply makes the definition of an electrically small antenna more stringent. However, examining the impedance properties of a number of electrically small wire antennas, it becomes evident that many small antennas

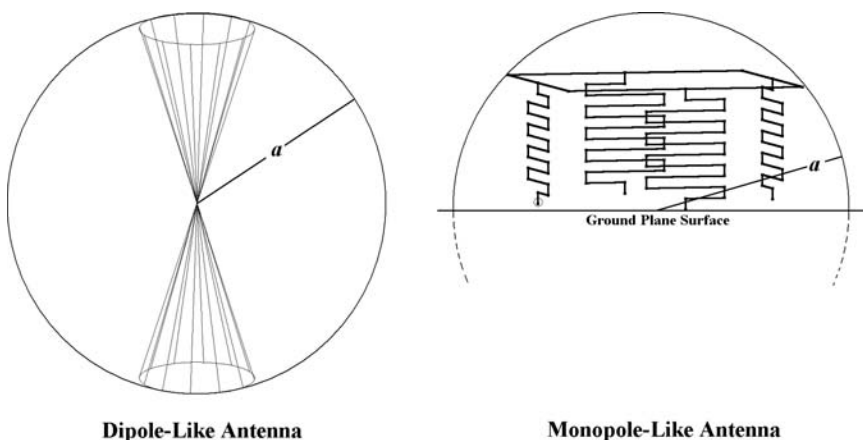


FIGURE 6-1 Definition of the radius a of an imaginary sphere circumscribing the maximum dimension of a free-space dipole-like antenna and monopole-like antenna mounted on a ground plane

exhibit similar behavior near the $ka = 0.5$ limit. As an example, the radiation resistances of certain classes of small dipole and monopole wire antennas converge to the value of the radiation resistance of the straight-wire dipole or monopole of the same height near $ka = 0.5$, independent of their total wire length or geometry.^{6,7} For these reasons, the definition of an antenna being electrically small is taken to be one where $ka \leq 0.5$.

6.2 THE ELECTRICALLY SMALL TRANSMIT ANTENNA

The feedpoint impedance of an antenna is defined as $Z_A(\omega) = R_A(\omega) + jX_A(\omega)$, where $R_A(\omega)$ is the frequency dependent feedpoint resistance, $X_A(\omega)$ is the frequency dependent feedpoint reactance, and ω is the radian frequency $2\pi f$, where f is the frequency in Hz. The frequency dependent resistance contains both radiation $[R_r(\omega)]$ and loss $[R_l(\omega)]$ terms from which the antenna's frequency dependent radiation efficiency $[\eta_r(\omega)]$ is determined using

$$\eta_r(\omega) = \frac{R_r(\omega)}{R_A(\omega)} = \frac{R_r(\omega)}{R_r(\omega) + R_l(\omega)} \quad (6-1)$$

The electrically small antenna behaves as either a lossy capacitor (C), a lossy inductor (L), or a combination of both, and its feedpoint impedance takes the form of a series or parallel RLC circuit. Because it behaves similarly to a lossy circuit element, it is often convenient to describe the properties of the small antenna in terms of circuit equivalent concepts, particularly the radiation power factor (P), quality factor (Q), fractional bandwidth (FBW), efficiency (η), and impedance matching. These will be discussed in later sections.

Consider the electrically small antenna connected to a transmitter or source through a lossless transmission line as illustrated in Figure 6-2. The source is represented by a Thevenin equivalent circuit having a voltage $V_s(\omega)$ and an impedance $Z_s(\omega)$ that are both assumed to be real and constant for all values of ω . The transmission line is assumed to have a real characteristic impedance $Z_{CH} = R_{CH}$. For simplicity in the transmission line and circuit analysis, it is also assumed that $R_{CH} = Z_s(\omega) = R_s$. The feedpoint impedance of the electrically small antenna is represented by a series RLC circuit (equivalent to a lossy capacitor) or by a parallel RLC circuit (equivalent to a lossy inductor).

The primary design objective when using the small antenna in the transmit mode is to radiate as much of the available source power as possible. The amount of power radiated by the small antenna is a function of two factors: the impedance match (or mismatch) between the antenna and the source (including the feeder transmission line), and the

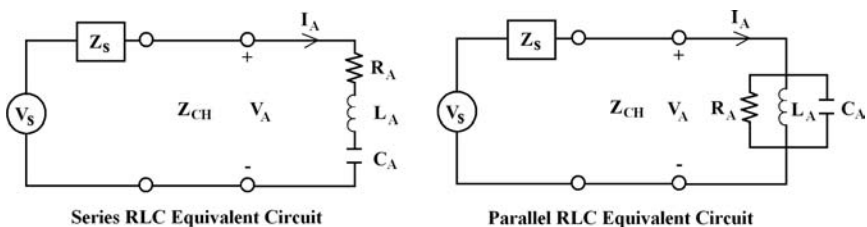


FIGURE 6-2 Series and parallel RLC equivalent circuits that can be used to represent the frequency dependent feedpoint impedance of the electrically small antenna

antenna's radiation efficiency. The level of impedance mismatch between the antenna and the source is characterized by using the antenna's VSWR, which is given by

$$\text{VSWR}(\omega) = \frac{1 + |\Gamma(\omega)|}{1 - |\Gamma(\omega)|} \quad (6-2)$$

where Γ is the usual reflection coefficient given by

$$\Gamma(\omega) = \frac{Z_A(\omega) - Z_{CH}}{Z_A(\omega) + Z_{CH}} \quad (6-3)$$

If the source drives the input of the lossless transmission line with a forward power, P_{FWD} , the total power incident at the antenna's feedpoint terminal is $P_{\text{INC}} = P_{\text{FWD}}$, and the total power accepted by (delivered to) the antenna, P_A , is given by

$$P_A = (1 - |\Gamma|^2) P_{\text{INC}} \quad (6-4)$$

The total power radiated by the antenna, P_r , is a function of the antenna's radiation efficiency and is given by

$$P_r = \eta_r P_A \quad (6-5)$$

Very often it is convenient to define the antenna's overall or realized efficiency, which is defined as the ratio of radiated power to incident power. The previously defined radiation efficiency of the antenna is the ratio of radiated power to accepted power. The realized efficiency of the antenna is given by

$$\eta_0(\omega) = \frac{P_r}{P_{\text{INC}}} = \eta_r(\omega) [1 - |\Gamma(\omega)|^2] \quad (6-6)$$

Radiating most of the available source power (P_{INC}) requires the small antenna to have high radiation efficiency and a reasonable impedance match to the source and feeder transmission line. The small antenna will radiate the same total power as any other antenna having the same values of radiation efficiency and VSWR, independent of its size relative to the operating wavelength. That said, achieving high efficiency and a good impedance match with an electrically small antenna can be challenging.

Finally it is important to note that the concept of designing the small antenna to match the impedance of free-space is not valid. All small antennas, regardless of size or geometry, are inherently matched to free-space. The only design issues to consider in determining the amount of radiated power are the antenna's radiation efficiency and VSWR: the small antenna must be impedance matched to the source.

6.3 THE ELECTRICALLY SMALL DIPOLE AND LOOP ANTENNAS

Prior to introducing the equivalent circuit concepts of radiation power factor, quality factor, fractional bandwidth, and impedance matching, the basic performance properties and characteristics of the electrically small dipole and loop antennas are described.

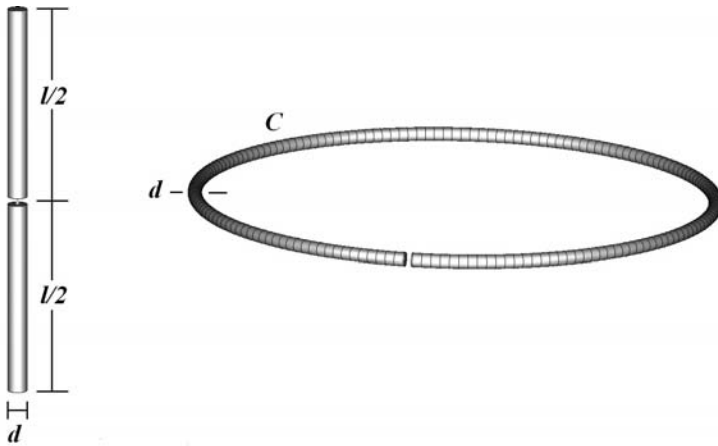
**Dipole****Loop**

FIGURE 6-3 Depictions of the straight-wire dipole and circular loop antennas. The straight-wire dipole has an overall length l and a conductor diameter d . The circular loop has a circumference C and a conductor diameter d .

These basic antenna elements are the foundation or starting point for the design of virtually all electrically small antennas. In some manner, all electrically small antenna designs can be shown to be derivatives of the simple straight-wire dipole, monopole, or loop antennas.*

The straight-wire dipole and circular loop antennas are depicted in Figure 6-3. The straight-wire dipole has an overall length l and a conductor diameter d . The circular loop has a circumference C and a conductor diameter d . The basic performance properties of interest include the feedpoint radiation and loss resistances, the feedpoint reactance, radiation efficiency, VSWR, directivity, gain, bandwidth, and quality factor. The bandwidth and quality factor of these antennas will be discussed in the next section.

The impedance properties of a straight-wire dipole antenna, having a length $l = 1$ m and a conductor diameter $d = 5$ mm ($l/d = 200$), are depicted in rectangular and Smith Chart format in Figure 6-4 over a frequency range of 5 through 150 MHz. The properties of the dipole are simulated using the NEC4 engine of EZNEC/4 Pro.⁸ Copper conductor loss is included in these results. The straight-wire dipole exhibits its first natural resonance $[X_A(\omega_0) = 0]$ at an approximate radian frequency $\omega_0 \approx c\pi/l$, where c is the speed of light in a vacuum. This first natural resonance is a series resonance, where the frequency derivative of the feedpoint reactance at the resonant frequency $[X'(\omega_0)]$ is greater than zero. The first natural resonant frequency for the dipole is approximately 142 MHz.

Below the first natural resonant frequency, the straight-wire dipole's feedpoint reactance is always negative (capacitive) and approaches $-\infty$ as $\omega \rightarrow 0$. The radiation resistance of the dipole approaches 0 as $\omega \rightarrow 0$. At DC, the dipole is an open circuit. At frequencies

* In some instances, the electrically small antenna can be shown to be a derivative of the basic slot radiator.

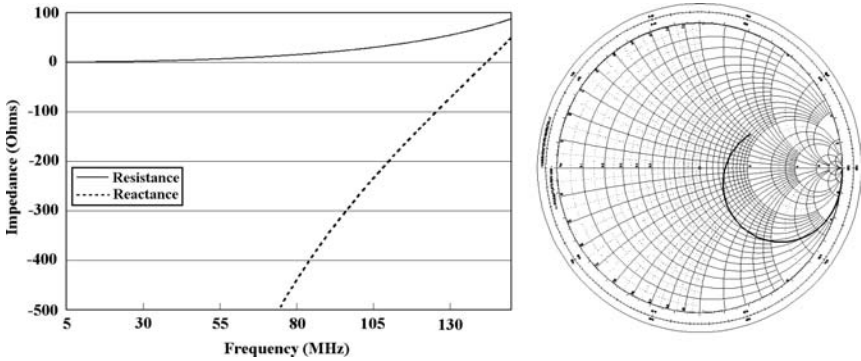


FIGURE 6-4 Feedpoint impedance of the straight-wire dipole antenna having an overall length of 1 m and a conductor diameter of 5 mm

near and below the small antenna limit, the radiation resistance of the straight-wire dipole is approximately given by⁹

$$R_r^d \approx 20\pi^2 \left(\frac{l}{\lambda} \right)^2 \quad (6-7)$$

which illustrates the radiation resistance's ω^2 behavior at low frequencies. The decreasing value of radiation resistance and the increasing value of capacitive reactance result in very high values of VSWR at low frequencies. The VSWR ($Z_{CH} = 50 \Omega$) of the 1-m dipole antenna as a function of frequency is presented in Figure 6-5. At values of $ka \leq 0.5$ ($f \leq 47.7$ MHz), the VSWR approaches values in excess of several thousand, making it impractical to deliver power to the dipole using a practical transmission line.

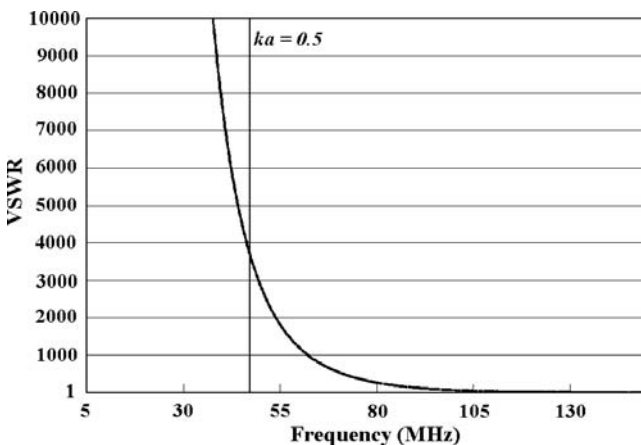


FIGURE 6-5 VSWR of the straight-wire dipole antenna having an overall length of 1 m and a conductor diameter of 5 mm

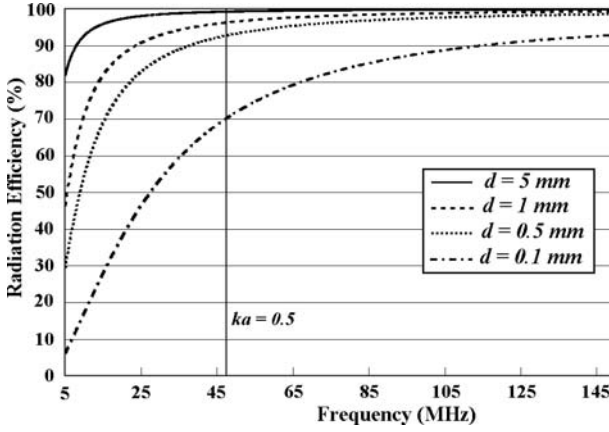


FIGURE 6-6 Radiation efficiency of the straight-wire dipole antenna having an overall length of 1 m and a conductor diameter d . Radiation efficiency curves are presented for several values of conductor diameter.

To calculate the radiation and realized efficiency of the small dipole, its loss resistance must be determined. If the skin depth δ in the conductor is somewhat less than $d/2$, the loss resistance of the dipole can be approximated as¹⁰

$$R_l^d = \frac{l\rho}{2\pi d\delta} \approx \frac{l}{2\pi d} \sqrt{\frac{kc\mu_0\rho}{2}} \quad (6-8)$$

where ρ is the resistivity of the conductor and μ_0 is the permeability of free-space. For δ somewhat greater than $d/2$, the loss resistance of the small dipole can be approximated by $R_l^d \approx 2l\rho/(\pi d^2)$. To obtain the approximation of loss resistance in Eq. 6-8, it is assumed that the current on the dipole has a cosine dependence over the length of the conductor and that the current density decays exponentially from its value at the surface of the conductor.

The radiation efficiency of the dipole versus frequency is presented in Figure 6-6 for several values of conductor diameter d . Well below values of $ka = 0.5$, the radiation efficiency of the small dipole remains in excess of 90 percent for conductor diameters larger than 0.5 mm. It is not until the value of ka is much less than 0.5 that the radiation efficiency of the small dipole diminishes to values where it rapidly approaches 0. This illustrates that reasonably high values of radiation efficiency can be achieved with an electrically small antenna provided the physical constraints allow the use of a suitable diameter conductor.*

In low-frequency applications ($f \leq 3000$ MHz), where electrically small antennas are more typically used, high radiation efficiencies can be achieved at small values of ka . In these applications, the overall or realized efficiency of the small antenna is diminished significantly by the mismatch loss associated with the antenna's high VSWR. From a practical perspective, the first significant design challenge in using an electrically small antenna is

* It is an important point to note here that the radiation efficiency of the dipole (or any small antenna) does not scale with ka (frequency). If the frequency, length, and diameter of the dipole (antenna) are simultaneously scaled, the radiation efficiency will not remain the same as a function of ka . This is addressed in more detail in Section 6.5.

improving its VSWR by matching the antenna to the feeder transmission line. Often the impedance matching process is itself inefficient, particularly if low Q (lossy) circuit elements are used to match the antenna. Losses within the matching network further reduce the realized efficiency of the small antenna.

A similar discussion on impedance and efficiency holds for the circular loop antenna, also depicted in Figure 6-3. A circular loop antenna, having a circumference C , exhibits its first natural resonance [$X_A(\omega_0) = 0$] at an approximate radian frequency $\omega_0 \approx c\pi/C$. This first natural resonance is a parallel resonance or antiresonance, where the frequency derivative of the feedpoint reactance at the antiresonant frequency [$X'(\omega_0)$] is less than zero. At its antiresonant frequency, the loop's resistance can be several thousand ohms, and the loop essentially behaves as an open circuit. All antennas, over some ranges of frequency, exhibit antiresonant behavior where $X'(\omega_0) < 0$ and the antenna does not obey the Foster reactance theorem.¹¹

At this point, several choices of loop circumference could be considered. A circular loop antenna with a circumference $C = 1$ m could be considered, in which case the conductor length would be the same as that of the small dipole. This loop would have a diameter equal to $1/\pi$ and would therefore occupy substantially less overall volume than the 1-m dipole. The first natural antiresonant frequency of this loop (≈ 140 MHz, for $d = 5$ mm) would nearly equal the first natural series resonant frequency of the dipole. The frequency at which $ka = 0.5$ (149.9 MHz) is approximately equal to the natural antiresonant frequency of the loop, substantially higher than the frequency at which $ka = 0.5$ for the dipole. For this reason, a loop having a diameter equal to 1 m is considered, so that the loop and dipole occupy the same overall volume and the frequencies at which $ka = 0.5$ for each are identical. The circumference of the loop is equal to 3.14 (π) m, and its first natural antiresonant frequency is approximately 45 MHz, which is substantially less than the first resonant frequency of the dipole as a result of its much longer conductor length.

The impedance of a lossy circular loop antenna having a circumference $C = 3.14$ m and a conductor diameter $d = 5$ mm is depicted in Figure 6-7 over a frequency range of 5 through 150 MHz. Below the first natural antiresonant frequency, the circular loop's feedpoint reactance is always positive (inductive) and approaches 0 as $\omega \rightarrow 0$. The radiation resistance of the loop also approaches 0 as $\omega \rightarrow 0$. At DC, the loop is a short circuit.

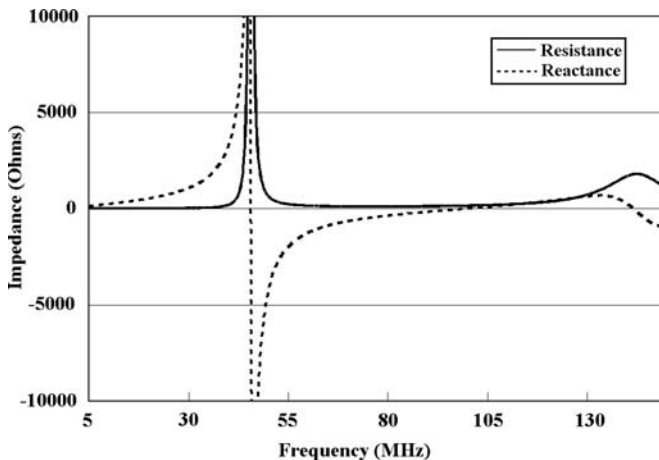


FIGURE 6-7 Impedance of the circular loop antenna having a circumference of 3.14 m and a conductor diameter of 5 mm

At frequencies near and below the small antenna limit ($ka \leq 0.5$), the radiation resistance of the loop is approximately given by⁹

$$R_r^l \approx 20\pi^2 N^2 \left(\frac{C}{\lambda} \right)^4 \quad (6-9)$$

where N is the number of total turns in the loop. The low-frequency behavior of the loop is substantially different from the low-frequency behavior of the dipole. While the dipole's radiation resistance exhibits an ω^2 behavior, the loop's radiation resistance exhibits an ω^4 behavior. This behavior has significant impact on the loop's efficiency and bandwidth properties at small values of ka . Similar to the dipole, the loop's decreasing value of radiation resistance results in very high values of VSWR at low frequencies, making it impractical to deliver power to the loop with typical transmission lines.

As with the dipole, calculating the radiation and realized efficiency of the small loop requires determining its loss resistance. If the skin depth δ in the conductor is somewhat less than $d/2$, the loss resistance of the circular loop can be approximated as¹⁰

$$R_l^l \approx \frac{NC\rho}{\pi d\delta} \approx \frac{NC}{\pi d} \sqrt{\frac{kc\mu_0\rho}{2}} \quad (6-10)$$

For δ somewhat greater than $d/2$, the loss resistance of the small circular loop can be approximated by $R_l^l \approx 4NC\rho/\pi\rho^2$. It is assumed that the current does not vary around the loop and the current density decays exponentially from its value at the surface of the conductor.

The radiation efficiency of the loop versus frequency is presented in Figure 6-8 for a conductor diameter of 5 mm. For values of ka at and just below 0.5 ($f \leq 47.7$ MHz), the radiation efficiency of the small loop remains in excess of 90 percent. However, with decreasing values of ka , the radiation efficiency of the loop diminishes and approaches 0 at a rate much quicker than that of the dipole due to the ω^4 dependence of the loop's radiation resistance.

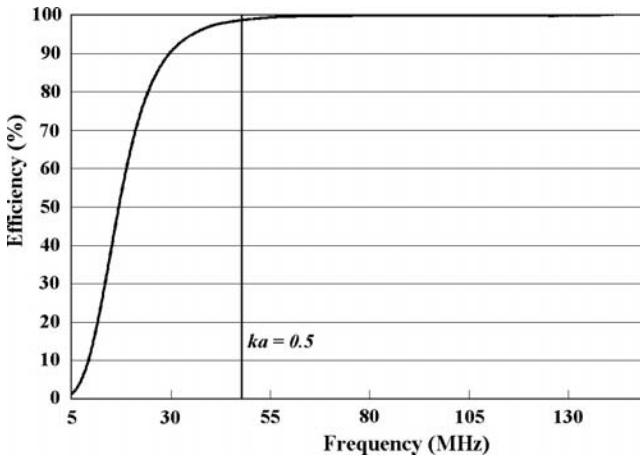


FIGURE 6-8 Radiation efficiency of the circular loop antenna having a circumference of 3.14 m (a diameter of 1 m) and a conductor diameter $d = 5$ mm

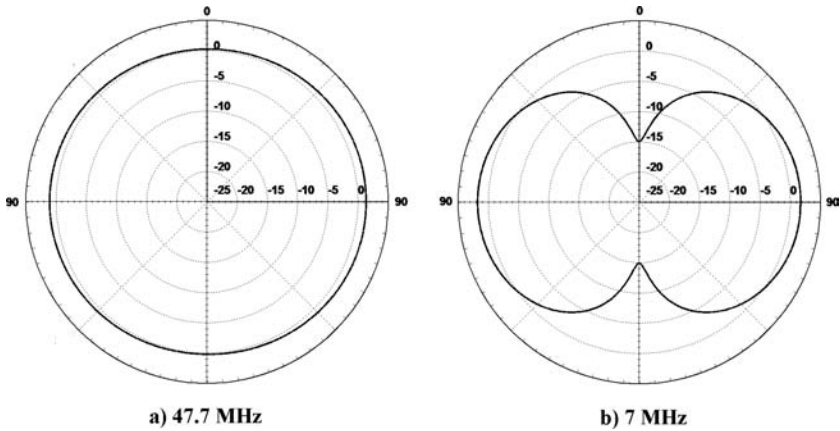


FIGURE 6-9 θ -sweep plane radiation pattern of the circular loop antenna having a circumference of 3.14 m (a diameter of 1 m) and a conductor diameter $d = 5$ mm at frequencies of (a) 47.7 MHz and (b) 7 MHz

At very small values of ka , a small loop is substantially less efficient than a dipole having the same conductor diameter and occupying the same overall volume. This difference in radiation efficiency for small values of ka causes the loop's quality factor and bandwidth properties to be substantially different from those of the small dipole. This will be discussed in detail in the next section.

A final point to note regarding the loop's radiation efficiency is that it can be increased by increasing the number of turns in the loop. As the number of turns increases, the loop's radiation resistance increases as N^2 while the loop's loss resistance increases as N .

The final performance properties of the small dipole and loop considered in this section are the directivity and radiation patterns, which help to describe the fundamental operating mode of the antenna. The straight-wire dipole antenna operates in a fundamental TM (electric dipole) mode from frequencies where $ka \ll 0.5$ through its first natural resonance. Throughout this entire range of frequencies, its radiation pattern is the well-known omnidirectional figure-eight. At its first natural resonant frequency, the directivity of the dipole is approximately 1.64 (2.14 dB).^{*} As the value of ka decreases, the directivity of the dipole approaches a constant value of approximately 1.5 (1.76 dB). For the 1-m dipole having a conductor diameter of 5 mm, the directivity approaches the 1.76-dB value at a frequency just below the frequency where $ka = 0.5$.

The small circular loop antenna is generally presumed to operate in a fundamental TE mode, where it behaves equivalently to a magnetic dipole. Unlike the electric dipole, however, the radiation pattern (mode) of the small loop changes significantly from frequencies where $ka \ll 0.5$ through frequencies where ka approaches 0.5, and the loop operates near its first natural antiresonance. For the loop to operate in a fundamental TE (magnetic dipole) mode, the current distribution around the loop must be nearly constant. This occurs in regions where $ka \ll 0.5$. The radiation pattern of the circular loop antenna at $ka = 0.5$ (47.7 MHz) is presented in Figure 6-9a. The loop lies in the xy plane, and the radiation

^{*} The directivity of the half-wavelength dipole (which is not naturally resonant) is slightly higher, having a value that approaches 2.18 dB.

pattern is an elevation (θ) sweep. At this frequency the loop does not exhibit the radiation pattern of a magnetic dipole, where well-developed nulls should exist in the $\pm z$ -axis directions. The directivity at this frequency is less than expected for a small antenna operating in its fundamental mode. The directivity is approximately 1.4 dB.

With decreasing values of ka , the current distribution on the loop becomes constant, and well-developed pattern nulls in the $\pm z$ -axis directions are formed. For the circular loop considered here, the TE (magnetic dipole) mode develops at frequencies where $ka < 0.1$ (≈ 10 MHz). The radiation pattern of the circular loop antenna at a frequency of 7 MHz is presented in Figure 6-9b. At this frequency the directivity approaches a constant value of approximately 1.5 (1.76 dB). The radiation efficiency of the loop in this frequency range is low, and the loop is an inefficient radiator.

6.4 POWER FACTOR, QUALITY FACTOR, AND BANDWIDTH

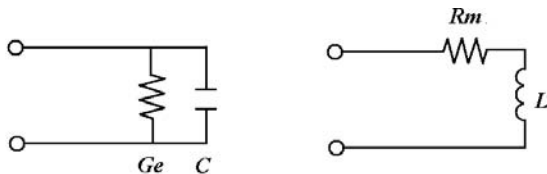
Wheeler introduced the concept of radiation power factor³ to quantify the radiation of real power from a small antenna that accepts a much larger value of reactive power. The power factor for small electric (dipole/capacitive) and magnetic (loop/inductive) antennas is defined in terms of equivalent circuit parameters (R , L , G , and C), depicted in Figure 6-10, and an equivalent cylindrical volume. Wheeler later provided further definition of radiation power factor in terms of the antenna's effective volume and in terms of the antenna's resistance and reactance properties. Wheeler also defined power factor as the inverse of the antenna's quality factor (Q).

The radiation power factor, P , is defined for electric and magnetic antennas in terms of their equivalent circuit parameters as

$$P_e = \frac{G_e}{\omega C}; \quad P_m = \frac{R_m}{\omega L} \quad (6-11)$$

where G_e , R_m , C , and L represent the parameters of the equivalent circuits depicted in Figure 6-10. Both P_e and P_m can be expressed in terms of the small antenna's feedpoint resistance and reactance as

$$P = \frac{R_A}{|X_A|} \quad (6-12)$$



Small Electric Antenna

Small Magnetic Antenna

FIGURE 6-10 Depictions of Wheeler's equivalent circuits for the small electric and magnetic antennas

For a lossless, electrically small antenna, both P_e and P_m exhibit an ω^3 dependence. The radiation power factor is also defined in terms of the spherical radius of the antenna's effective volume, a' , as

$$P = \frac{2}{9} (ka')^3 \quad (6-13)$$

which also exhibits an ω^3 dependence. The maximum power factor attainable for any antenna occupying a physical volume of spherical radius a is given by

$$P_{\max} = (ka)^3 \quad (6-14)$$

which implies that the maximum achievable effective volume is $1.651a$ for an antenna occupying a physical volume of spherical radius a . The significance of Eq. 6-14 is that it establishes an upper bound on the achievable power factor as a function of the antenna's size relative to the operating wavelength. Given the inverse relation between power and quality factor, Eq. 6-14 also establishes a lower bound on Q as a function of the antenna's size.

The Q and bandwidth are defined for the tuned or self-resonant small antenna. A tuned antenna is one whose feedpoint reactance is tuned to zero at a frequency ω_0 using a lossless, single component (inductor or capacitor), series tuning circuit, having a reactance of $X_S(\omega_0) = -X_A(\omega_0)$. The tuned antenna's impedance is defined as $Z_0(\omega) = R_0(\omega) + jX_0(\omega) = R_A(\omega) + j[X_S(\omega) + X_A(\omega)]$ and $Z_0(\omega_0) = R_A(\omega_0)$. The quality factor of the tuned antenna, $Q(\omega_0)$, is defined in terms of the ratio of internal energy, $W(\omega_0)$, and accepted power, $P_A(\omega_0)$, as^{12,13}

$$Q(\omega_0) = \frac{\omega_0 |W(\omega_0)|}{P_A(\omega_0)} \quad (6-15)$$

which has been shown to be approximately equal to¹³

$$Q(\omega_0) \approx Q_Z(\omega_0) = \frac{\omega_0}{2R_A(\omega_0)} \sqrt{R_A'(\omega_0)^2 + \left(X_A'(\omega_0) + \frac{|X_A(\omega_0)|}{\omega_0} \right)^2} \quad (6-16)$$

In a communication system, engineers are generally more concerned with defining and characterizing the antenna's operating bandwidth. This might lead you to question the necessity of characterizing the antenna's Q . The Q of the small antenna is characterized for two significant reasons. First, the bandwidth of an antenna is often ambiguously defined. Bandwidth may be defined in terms of the conductance bandwidth or the VSWR bandwidth for an arbitrary choice of VSWR and an arbitrary choice of feeder transmission-line characteristic impedance, Z_{CH} . Second, and more importantly, fundamental limits on the minimum achievable Q (the lower bound) can be defined in terms of the antenna's size relative to the operating wavelength (ka). Characterizing the antenna's Q allows you to quantify how well the antenna performs relative to theoretical limits. Third, if an inverse relationship between bandwidth and Q is established, a fundamental limit on the maximum achievable bandwidth in terms of the antenna's size relative to the operating wavelength can be defined.

For these reasons, it is necessary to define a bandwidth that is related to the inverse of Q over a wide range of frequencies. A suitable definition of bandwidth for this purpose

is matched VSWR bandwidth,* $\text{FBW}_V(\omega_0)$, where the VSWR (denoted by s) of the tuned antenna is determined under the condition that the characteristic impedance, Z_{CH} , of the transmission line connecting the antenna to the matched source is equal to the antenna's feedpoint resistance at the tuned frequency: $Z_{\text{CH}} = R_a(\omega_0)$. Matched VSWR bandwidth is equal to

$$\text{FBW}_V(\omega_0) = \frac{\omega_+ - \omega_-}{\omega_0} \quad (6-17)$$

where ω_+ and ω_- are the frequencies above and below ω_0 , respectively, where the VSWR is equal to s . Matched VSWR bandwidth and Q are related as¹³

$$Q(\omega_0) = \frac{2\sqrt{\beta}}{\text{FBW}_V(\omega_0)}, \quad \sqrt{\beta} = \frac{s-1}{2\sqrt{s}} \leq 1 \quad (6-18)$$

The approximations in Eq. 6-16 and Eq. 6-18 were derived in Yaghjian and Best¹³ under the assumptions that the half-power matched VSWR bandwidth ($s = 5.828:1$) of the antenna is not too large and that the antenna exhibits a single resonance within its defined operating bandwidth. In some instances where the antenna exhibits multiple, very closely spaced resonances within its operating bandwidth and $R_A'(\omega)$ and $X_A'(\omega)$ approach 0, these approximations may not hold.

The lower bound on Q , often referred to as the *Chu limit*, is given by^{12,14}

$$Q_{lb} = \eta_r \left(\frac{1}{(ka)^3} + \frac{1}{ka} \right) \quad (6-19)$$

For values of ka less than 0.5 (0.34), Eq. 6-19 is within 20 percent (10 percent) of the inverse of the upper bound on radiation power factor stated in Eq. 6-14. A significant point to note is that the Q of an electrically small antenna is often incorrectly assumed or stated to be equal to this lower bound. The Q of all electrically small antennas will be greater than this lower bound.

From Eq. 6-18 and Eq. 6-19, an upper bound on the maximum achievable matched VSWR bandwidth can be defined in terms of the antenna's size relative to the operating wavelength (ka). This upper bound is given by

$$\text{FBW}_{\text{vub}} = \frac{1}{\eta_r} \frac{2\sqrt{\beta}(ka)^3}{1+(ka)^2} = \frac{1}{\eta_r} \frac{(ka)^3}{1+(ka)^2} \frac{s-1}{\sqrt{s}} \quad (6-20)$$

This upper bound is defined at the antenna's feedpoint terminal and does not account for any realized increase in bandwidth that may occur through the use of external matching components. Additionally this upper bound is defined for a matched antenna ($s = 1$ at ω_0). It is well known that an increase in realized bandwidth can be achieved if the antenna is not perfectly matched at the tuned or self-resonant frequency.

* Often conductance bandwidth is related to the inverse of Q . *Conductance bandwidth* is defined as the accepted power bandwidth (typically the half-power bandwidth) of the antenna under the condition that the antenna's feedpoint voltage is held constant versus frequency. This definition of bandwidth is valid provided that the tuned antenna operates in a resonance region where $X_A'(\omega_0) > 0$. If the tuned antenna is operated in an antiresonance region where $X_A'(\omega_0) < 0$, the conductance bandwidth cannot be defined. The matched VSWR bandwidth can be defined in all ranges of frequency, including regions of both resonance and antiresonance.

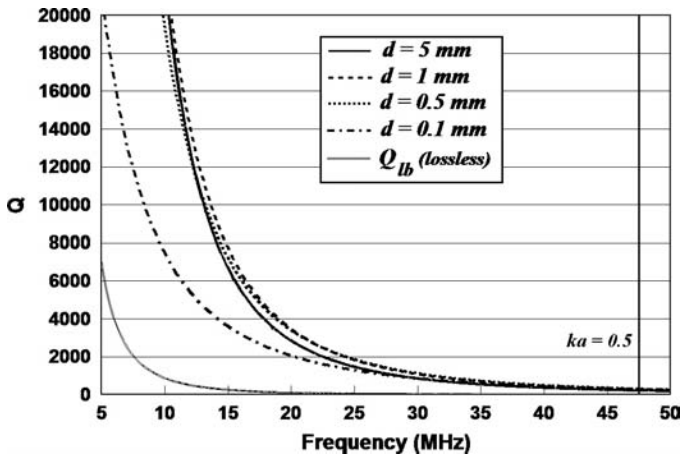


FIGURE 6-11 Q of the small, lossy, straight-wire dipole antenna and the lower bound on Q for the lossless dipole. The dipole has a length $l = 1$ m and a conductor diameter $d = 5$ mm.

The Q of the small dipole ($l = 1$ m) is determined using Eq. 6-16 for the various values of conductor diameter d : 5 mm, 1 mm, 0.5 mm, and 0.1 mm. The Q 's of the lossy dipoles are compared with the lower bound on Q in Figure 6-11. The lower bound is presented for the lossless dipole since the lower bound for the lossy dipoles differs as a function of the dipoles' radiation efficiencies. From Figure 6-11, the substantial difference between the dipole's actual Q 's and the lower bound is evident. The Q 's of all the dipoles are substantially higher than the lower bound. Note that the lower bound for the lossy antennas is less than the lower bound for the lossless antenna. While the lower bound exhibits a $1/\omega^3$ frequency dependence, the actual Q 's of the small dipoles exhibit a $1/\omega^4$ or $1/\omega^5$ dependence. Similar behavior is also exhibited by the lossless dipole.

Since electrically small antennas exhibit behavior that can be generally described in terms of equivalent circuit elements, the Q of the small antenna is often approximated by the value of $|X_A|/R_A$. In Figure 6-12 the Q of the 5-mm-diameter dipole is compared with the corresponding values of $|X_A|/R_A$ as a function of frequency. For $ka < 0.5$, where Q is typically very large, the value of $|X_A|/R_A$ provides a reasonable approximation of the antenna Q . At $ka = 0.5$, the value of Q is approximately 206 and the value of $|X_A|/R_A$ is approximately 187. With increasing values of ka and decreasing value of Q , the validity of the approximation degrades significantly. This is also true for electrically small antennas that are designed to be self-resonant where $X_A(\omega_0) = 0$ and $|X_A|/R_A = 0$, implying a zero Q , which is not a valid result. There is no such thing as a zero Q electrically small antenna. The values of $|X_A|/R_A$ must be used judiciously in approximating the value of antenna Q .

The behavior of the lossy loop antenna is substantially different from that of the lossy dipole antenna at small values of ka . The difference in their relative behavior is a function of the difference in the two antennas' radiation efficiencies. This difference is a result of the ω^2 frequency dependence of the small dipole's radiation resistance and the ω^4 frequency dependence of the small loop's radiation resistance. The loss resistances of both the small dipole and loop are proportional to the conductor length, inversely proportional to the conductor diameter, and exhibit an $\omega^{1/2}$ dependence.

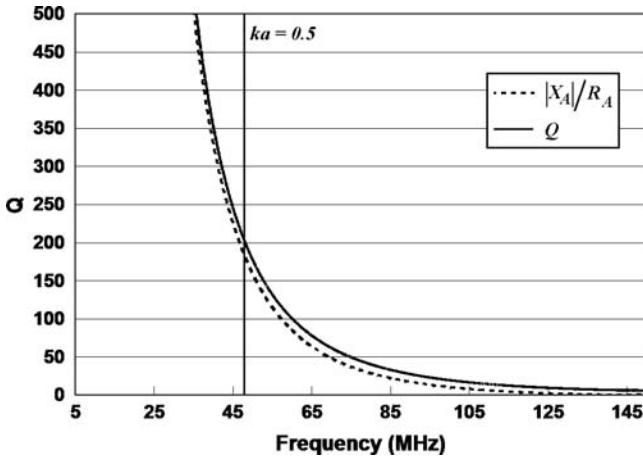


FIGURE 6-12 A comparison of the small dipole's Q and the value of $|X_A|/R_A$. The dipole has a length $l = 1$ m and a conductor diameter $d = 5$ mm.

The Q of the lossy circular loop ($C = 3.14$ m, $d = 5$ mm) is compared with the corresponding values of $|X_A|/R_A$ and the lower bound on Q in Figure 6-13. In this instance, the lower bound is presented for the lossy antenna. In frequency ranges where $ka \ll 0.5$ and the loop behaves as a magnetic dipole, the substantially decreasing radiation efficiency of the loop dominates its performance. At low frequencies, the Q of the loop is substantially

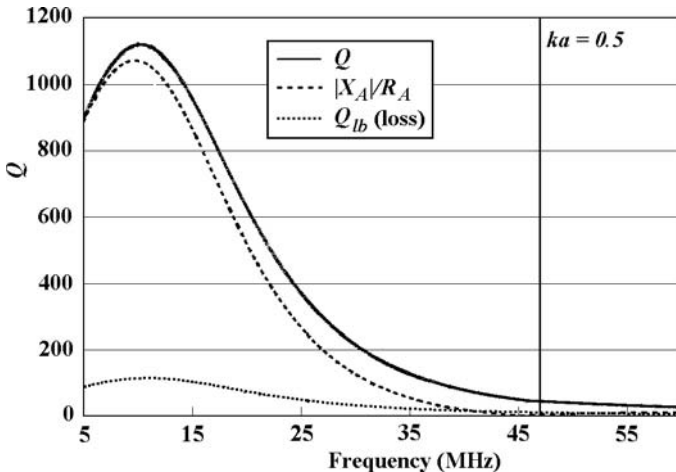


FIGURE 6-13 A comparison of the small loop's Q and the value of $|X_A|/R_A$. The lower bound on Q for the lossy loop is also presented. The loop has a circumference $C = 3.14$ m and a conductor diameter $d = 5$ mm.

larger than the lower bound, but neither obeys a $1/\omega^n$ dependence.* You must be cautious in measuring the bandwidth of the very small loop ($ka \ll 0.5$) and interpreting its relation to the lower bound on Q . The bandwidth of the very small loop can increase with decreasing frequency, which is consistent with the lower bound for the lossless loop as is evident in Figure 6-13. In this frequency region the increase in bandwidth is coupled with a substantial decrease in radiation efficiency.

6.5 RADIATION EFFICIENCY AND REALIZED GAIN

Considering that there is a fundamental lower bound on the minimum achievable Q as a function of antenna size relative to the operating wavelength, you might assume that there is also a fundamental limit on the maximum realized gain that can be achieved as a function of antenna size. In fact, there is not, and the maximum theoretical gain that can be achieved by the electrically small antenna is its directivity. In practice, the antenna's realized gain is of course limited by the antenna's radiation and matching efficiencies. Any small antenna can be impedance matched to the source at any single frequency. The bandwidth of the impedance match is defined by the inverse of Q and is generally limited by the frequency dependent VSWR (mismatch loss) rather than the frequency dependent radiation efficiency. Finally, if the antenna is impedance matched using an external matching network, the ohmic losses within the matching circuit must be included in the determination of the antenna's realized gain.

From Figure 6-6, it is evident that the small dipole can exhibit very high radiation efficiencies given an appropriate choice of conductor diameter. At $ka \approx 0.229$ (21.85 MHz), the radiation efficiency of the dipole is approximately 97.6 percent for a conductor diameter of 5 mm. However, the impedance of the dipole at this frequency is $1.02 - j2209 \Omega$, and the dipole's realized gain in a $50\text{-}\Omega$ system is approximately -42 dBi , which is primarily a result of the very high mismatch loss.

In previous editions of this *Handbook*, Wheeler discussed the relationship between radiation power factor and antenna gain. He stated that the operating efficiency (the realized gain) of the small antenna is limited by its radiation power factor (or the inverse of Q). He also stated that in general, greater size yields greater efficiency and that a specified efficiency imposes a requirement of some size. These statements must be interpreted judiciously in that you should not conclude that there is a fundamental relationship between realized gain and the antenna's power factor (or the inverse of Q) and size (ka). As the value of ka decreases and there is a corresponding decrease in the maximum achievable power factor and minimum achievable Q , there is not a corresponding decrease in the maximum achievable realized gain.

In stating that the operating efficiency of the small antenna is limited by its radiation power factor, Wheeler assumed the presence of an external matching network that would be necessary to match the small antenna to some desired characteristic impedance. With a small value of radiation power factor, the small antenna would have a low resistance and correspondingly much higher reactance. When matching a small antenna with a lossy network, greater matching efficiency can be achieved by increasing the value of the antenna's resistance and decreasing the value of the antenna's reactance: a higher power factor. In this context, greater operating efficiency can be achieved with a higher radiation power factor. However, concluding that a direct relation between realized gain and power factor

* The Q of the lossless loop will always increase with decreasing frequency, and it will obey a $1/\omega^n$ dependence.

or Q holds in all cases or that power factor, Q , or antenna size (ka) limits the achievable radiation efficiency or realized gain would be a mistake.

Reconsider the dipole antenna having a realized gain of -42 dBi at 21.85 MHz ($ka \approx 0.229$). At this frequency, the dipole's Q is 2195, which is approximately 25 times the lower bound of 87.7. While there is substantial margin for improving the antenna's Q , only marginal improvement can be achieved in terms of increasing the dipole's radiation efficiency. Very often, techniques used to lower the antenna Q result in a decrease in radiation efficiency. Similarly, an increase in radiation power factor does not always directly correspond to an increase in radiation efficiency. The radiation power factor of the dipole is approximately 4.64×10^{-4} .

The small dipole can be impedance matched to 50Ω using an external matching network comprised of a series inductor and a parallel inductor. Assuming that these lossy matching inductors have Q 's of 150,^{*} the values of inductances necessary to match the dipole to 50Ω ($s = 1.043$) are series inductance = $15.915 \mu\text{H}$ and parallel inductance = $0.251 \mu\text{H}$. The total matching efficiency of this network (reflection + loss) is approximately 6.48 percent, and the realized gain of the antenna is increased from -42 dBi to -10.27 dBi.

The question to consider is how much can the realized gain of the small antenna be increased through judicious antenna design. With the series and parallel inductor matching circuit, there is substantial ohmic loss within the series inductor that is used to make the small dipole nearly resonant.^{**} The same matching effect achieved with the lumped series inductor at the feedpoint can be implemented by adding more conductor length to the dipole and then winding it into a coil so as to maintain the same overall antenna length, l . This configuration is the well-known normal mode helix antenna.^{4,15} Incorporating series inductance within the antenna structure is generally more efficient than using a lumped series matching inductor at the feedpoint.

An example of a normal mode helix antenna, designed to be self-resonant near 21.5 MHz, is depicted in Figure 6-14. This normal mode helix has an overall length of 1 m and is self-resonant at a frequency of 21.46 MHz with a total feedpoint resistance of 2.87Ω and a radiation efficiency of 79.1 percent. The conductor diameter is 5 mm. This antenna can be impedance matched to 50Ω using a parallel inductor ($L = 0.091 \mu\text{H}$; $Q = 150$) at the feedpoint, which reduces the realized efficiency and gain to 76.96 percent and $+0.63$ dBi, respectively. The match frequency is 21.37 MHz, slightly lower than the resonant frequency of 21.46 MHz. The gain of the antenna is within 1.16 dB of its directivity at $ka = 0.224$.

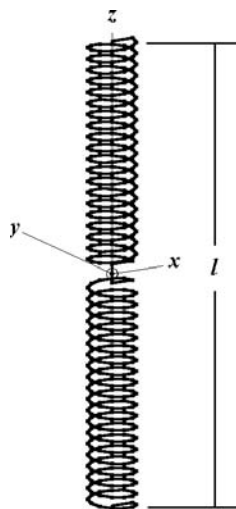


FIGURE 6-14 Depiction of the normal mode helix antenna having an overall length of 1 m and a conductor diameter of 5 mm. The resonant frequency of the antenna is 21.46 MHz.

^{*} For lumped circuit elements, Q is primarily used to characterize the efficiency of the element. The Q of the lumped circuit element is given by the ratio of the element's reactance to its resistance, $|X|/R$, where R represents total ohmic loss. A high Q is desirable since it indicates that the circuit element has low ohmic loss and is more efficient. For antennas, the resistance (R_A) represents the total power dissipation associated with both radiation and loss. A high radiation resistance is acceptable, and a low Q is desirable since it implies high bandwidth. For the lumped circuit element, a lower Q also indicates relatively wider bandwidth. The trade-off between bandwidth and loss is relevant for both the lumped circuit element and antenna.

^{**} In using a series parallel inductor combination to match a highly capacitive small antenna, the series inductor is used to reduce (tune) the high feedpoint capacitance at a frequency where the antenna is nearly resonant (the impedance has a small value of capacitive reactance). The parallel inductor is then used to transform the impedance to 50Ω .

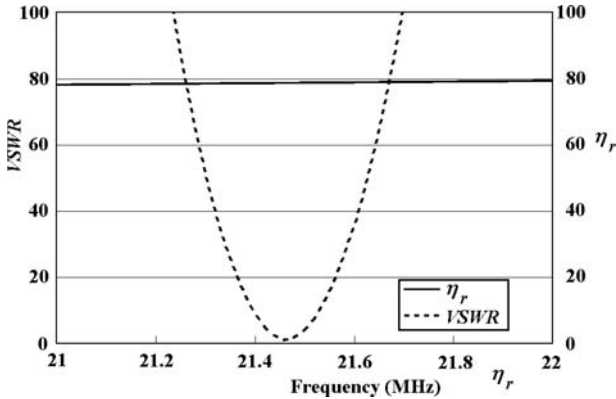


FIGURE 6-15 The VSWR and radiation efficiency of the small normal mode helix antenna. The VSWR is calculated assuming that the characteristic impedance of the feeder transmission line is equal to the antenna's resistance at the resonant frequency of 21.46 MHz.

The same normal mode helix is also inherently matched to a characteristic impedance of $2.87\ \Omega$ at its self-resonant frequency. In a $2.87\text{-}\Omega$ system, the antenna does not require further impedance matching, and it has a realized gain of approximately $+0.75\text{ dBi}$. A comparison of VSWR (mismatch loss) and radiation efficiency is presented in Figure 6-15. The radiation efficiency maintains a near constant value over a bandwidth much wider than that defined by the mismatch loss. The operating bandwidth of the small antenna is generally limited by the antenna's mismatch loss rather than the antenna's radiation efficiency.

Also note that the realized gain of the small antenna (the mismatch loss component) can be arbitrarily changed with selection of the characteristic impedance. The power factor and Q of the antenna are both independent of the choice of characteristic impedance, and a direct relationship between the antenna's realized gain and Q cannot be established.

The Q of the normal mode helix at its resonant frequency (21.46 MHz) is approximately 462, which is approximately five times the lower bound of 92. With the normal mode helix design, a substantially lower Q is achieved, but there is a decrease in radiation efficiency relative to the straight-wire dipole. The improvement in Q (or power factor) did not translate into an improvement in radiation efficiency.

Wheeler defined both a radiation power factor (radiation Q) and a loss power factor (loss Q). The radiation and loss power factors of the small dipole at 21.85 MHz are 4.52×10^{-4} and 1.12×10^{-5} , respectively (these values are determined from $R/|X|$). The radiation efficiency of the small dipole is 97.6 percent. The radiation and loss power factors of the small normal mode helix at 21.46 MHz are 1.17×10^{-3} and 0.46×10^{-3} , respectively (these values are determined from the inverse of Q since the value of $R/|X|$ at the self-resonant frequency is undefined). The radiation efficiency of the small normal mode helix is 79.1 percent. The substantial increase in radiation power factor of the normal mode helix relative to the dipole did not translate into an increase in radiation efficiency.

The final point of discussion regarding the radiation efficiency of the small antenna is its behavior as a function of scaling the size of the small antenna with changes in frequency. The radiation efficiency of the small antenna is a function of its radiation and loss resistances as defined in Eq. 6-1. Equations 6-7 and 6-9 show that the radiation resistances of the small dipole and loop scale with l/λ and C/λ , respectively. This is an expected result—if the antenna dimensions are doubled and the frequency is halved, maintaining the same value of ka , the antenna's radiation resistance is expected to be the same. The same behavior

does not hold for the loss resistance of the small antenna. Equations 6-8 and 6-10 show that the loss resistance does not scale directly with frequency but rather changes as a function of ω^2 . For this reason, the radiation efficiency of the small antenna does not scale with frequency or ka .

For example, the small normal mode helix antenna ($l = 1$ m; $d = 5$ mm) has a radiation efficiency of 79.1 percent at $ka \approx 0.224$ (21.46 MHz). Scaling the antenna dimensions by a factor of 10 and lowering the operating frequency to 2.146 MHz (maintaining the same value of ka) results in the antenna having a radiation efficiency of 92.5 percent. Scaling the antenna dimensions by a factor of 0.01 and increasing the operating frequency to 2146 MHz results in the antenna having a radiation efficiency of 26.6 percent. If the conductor diameter is not scaled with frequency but rather held constant, the opposite effect is observed. Maintaining a constant diameter with increasing frequency can become impractical since the antenna's volume decreases and the use of large diameter conductors is not possible. Often it is more practical to achieve higher radiation efficiency at lower frequencies where the use of larger diameter conductors is more practical.

6.6 THE ELECTRICALLY SMALL RECEIVE ANTENNA

In the transmit mode, where the antenna acts as a load, the primary objective is to have the small antenna radiate as much of the available source power as possible. In the receive mode, where the antenna acts as a source, the primary objective is to deliver as much signal power as possible to the receiver while maintaining a necessary level of signal-to-noise ratio. The small receive antenna can be modeled using a Thevenin equivalent circuit, where the antenna is the source having a voltage equal to the open-circuit voltage developed at the antenna's feedpoint terminal given an incident electromagnetic wave, and having a source impedance equal to the antenna's impedance $R_A + jX_A$, which is the same as the antenna's impedance in the transmit mode.

The equivalent circuit for the small receive antenna is depicted in Figure 6-16. $V_{oc}(\omega)$ is the antenna's open-circuit voltage established by an incident electromagnetic (EM) wave having a field intensity of E (V/m).^{*} $Z_A(\omega)$ is the antenna's frequency dependent feedpoint impedance, and $Z_L(\omega) = R_L(\omega) + jX_L(\omega)$ is the impedance seen at the antenna's feedpoint looking toward the receiver, which in practice is usually a constant real value such as $50\ \Omega$. $I_L(\omega)$ and $V_L(\omega)$ are the current and voltage developed at the equivalent load impedance $Z_L(\omega)$.

Often engineers struggle with the concept of relating the antenna's physical size to its ability to capture or extract power from an incident EM wave. Other than its relationship in establishing the antenna's radiation efficiency and the impedance match to the receiver, the physical size of the antenna is not relevant to its ability for capturing power from a passing EM wave. All antennas capture RF power (P_c) from an incident EM wave as a function of

$$P_c = \frac{|E|^2}{240\pi} \eta_r (1 - |\Gamma|^2) A_e = \frac{|E|^2}{240\pi} \eta_r (1 - |\Gamma|^2) \frac{\lambda^2 D}{4\pi} \quad (6-21)$$

^{*} In the formulations that follow, we assume that $|E|$ represents a *peak* value of voltage not an *rms* value. The same is true for the current and voltage used in any circuit analysis that follows from Figure 6-16. Power density is then determined using $\frac{1}{2} |E|^2 / (120\pi)$, and power is determined using $\frac{1}{2} |I|^2 R$.

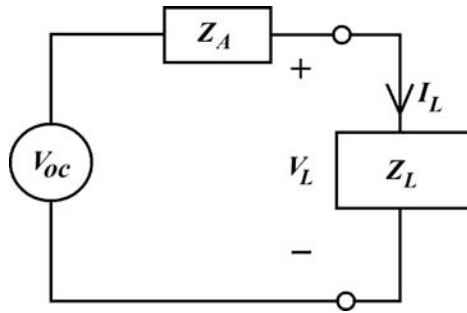


FIGURE 6-16 The Thevenin equivalent circuit of the small antenna operating in the receive mode

where P_C is the power delivered to the load impedance connected at the antenna's feedpoint terminal, Γ is the usual antenna reflection coefficient, A_e is the lossless antenna's effective area, and D is the antenna's directivity. This formula for captured power is stated with the assumption that the antenna and the incident EM wave are polarization matched, and that the load impedance connected at the antenna's terminal is a pure resistance and equal to the value of Z_{CH} used to calculate Γ . If the load connected at the antenna's terminal is complex, Eq. 6-21 becomes^{*16}

$$P_C = \frac{|E|^2}{240\pi} \eta_r \frac{(1 - |\Gamma|^2)(1 - |\Gamma_L|^2)}{|1 - \Gamma\Gamma_L|^2} \frac{\lambda^2 D}{4\pi} \quad (6-22)$$

where Γ_L is the reflection coefficient for the load impedance, Z_L . The calculation of the reflection coefficients in Eq. 6-22 is made with an arbitrary choice of real characteristic impedance, Z_{CH} . However, both must be calculated using the same value of Z_{CH} . The significance of Eq. 6-21 and Eq. 6-22 is that they indicate that the difference in captured power between a small antenna and a much larger antenna is a function of the differences in radiation efficiency, impedance match, and directivity.

To illustrate these concepts, consider two dipole antennas, one small, the other nearly $\lambda/2$ in length, where both are efficient and impedance matched to 50 Ω . The near- $\lambda/2$ antenna is a dipole with additional conductor length added at the upper and lower ends of the dipole to decrease the resonant frequency to a point where the resonant resistance is 50 Ω . The small antenna is a four-arm folded spherical helix, which is inherently matched to 50 Ω . Both antennas are depicted in Figure 6-17. The folded spherical helix antenna has an overall length of 79.16 cm and a conductor diameter of 10.71 mm. The dipole antenna has an overall length of 326.6 cm and a conductor diameter of 8.17 mm. Both antennas operate at a frequency of 29.97 MHz.

It is assumed that an EM plane wave having a field intensity of 100 V/m is incident at each antenna. The power delivered to a 50- Ω load is determined using NEC4 and Eq. 6-21. The NEC4 antenna simulations include the effects of cooper loss. Results for both antennas

^{*} With all antennas, maximum possible power capture occurs when the antenna is lossless and the load connected at the antenna feedpoint terminal is the conjugate of the antenna impedance ($Z_L = Z_A^*$). If the load impedance is complex, the calculation of the normal antenna reflection coefficient, Γ , does not provide a correct indication of the power transfer to the load and Eq. 6-21 cannot be used to calculate power capture.

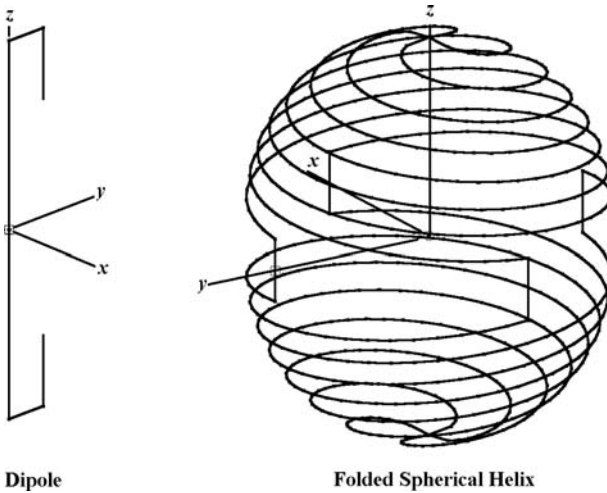


FIGURE 6-17 Depictions of the short dipole and electrically small folded spherical helix antennas. The short dipole has an overall length of 326.6 cm, and the folded spherical helix has an overall length of 79.16 cm. Both antennas are self-resonant at a frequency of 29.97 MHz with a VSWR less than 1.1 on a 50- Ω characteristic impedance.

and a self-resonant straight-wire dipole are presented in Table 6-1. The straight-wire dipole has an overall length of 478.8 cm and a conductor diameter of 9.57 mm. It is evident that all of the antennas capture nearly equal values of power from the incident EM wave. The difference in captured powers is primarily attributable to the difference in the antennas' directivities.

The amount of power captured by the small antenna can also be determined using simple circuit analysis if the open-circuit voltage at the antenna's feedpoint terminal is known. From the circuit of Figure 6-16, the captured power is given by

$$P_C = \frac{1}{2} |I_L|^2 R_L \quad (6-23)$$

TABLE 6-1 Comparison of Captured (Load) Power for Small Folded Helix Antenna and Two Larger Dipole Antennas (Frequency = 29.97 MHz; Incident Field Intensity = 100 V/m)

Antenna	D (dB)	Z_A (Ω)	η_r (%)	Z_L (Ω)	P_C (Watts) (NEC4)	P_C (Watts) (6-21)
Straight-wire dipole	2.14	$72.2 + j0.99$	99.8	72.2	172.7	172.51
Short dipole	1.93	$49.12 - j1.38$	99.8	50	164.05	164.33
Folded spherical helix	1.75	$46.62 - j0.22$	98.2	50	155.6	154.98

where $|I_L| = |V_{oc}/(Z_A + Z_L)|$. The open-circuit voltage developed at the feedpoint terminal of any antenna* can be determined given the incident field intensity. The receive antenna acts as a constant power source where the open-circuit voltage developed at the feedpoint is a value such that the power delivered to a conjugate matched load is given by

$$P_{CM} = \frac{|E|^2}{240\pi} \eta_r \frac{\lambda^2 D}{4\pi} = \frac{1}{2} |I_L|^2 R_A = \frac{1}{8} \frac{|V_{oc}|^2}{R_A} \quad (6-24)$$

The open-circuit voltage is then given by

$$|V_{oc}| = |E| \left| \frac{\lambda}{\pi} \sqrt{\frac{R_r D}{120}} \right| \quad (6-25)$$

Since the open-circuit voltage can be stated as $|V_{oc}| = |E| l_e$, where l_e is the effective length of the antenna, it follows that

$$l_e = \frac{\lambda}{\pi} \sqrt{\frac{R_r D}{120}} \quad (6-26)$$

For the small dipole of length l oriented along the z -axis, having a radiation resistance given by Eq. 6-7 and a peak or maximum directivity of 1.5 that occurs at $\theta = 90^\circ$, the effective length is given by

$$l_e = \frac{l}{2} \sin \theta \quad (6-27)$$

For the small loop ($ka \ll 0.5$) of circumference C lying in the xy plane, having a radiation resistance given by Eq. 6-9 and a maximum directivity of 1.5 that occurs at $\theta = 90^\circ$, the effective length is given by

$$l_e = \frac{NC^2}{2\lambda} \sin \theta \quad (6-28)$$

A comparison of the open-circuit voltages developed at the feedpoints of the small dipole and loop antennas described in Section 6.3 is presented in Table 6-2. The open-circuit voltage is determined at a frequency of 10 MHz ($ka \ll 0.5$), assuming an incident field strength of 10 V/m. Theoretical calculations using Eq. 6-27 and Eq. 6-28 are compared with NEC4 simulations.

Electrically small loops are often used as receive antennas. If the small dipole and loop are oriented to exhibit the same polarization, have the same radiation efficiency, the same directivity, and are connected to a conjugate

TABLE 6-2 Comparison of Open-Circuit Voltages Developed at Feedpoint Terminals of Small Dipole and Loop Antennas

Antenna	$V_{oc} (V) E l_e$	$V_{oc} (V) \text{NEC4}$
Small dipole	5	4.864
Small loop	1.646	1.764

The dipole has an overall length of 1 m and a conductor diameter of 5 mm. The loop has an overall diameter of 1 m and a conductor diameter of 5 mm. The incident field strength is 10 V/m. The frequency is 10 MHz. $\theta = 90^\circ$.

* The antenna is assumed to have a single feedpoint terminal.

TABLE 6-3 Comparison of Scattering Cross-Section for Small Folded Helix Antenna and Two Larger Dipole Antennas (Frequency = 29.97 MHz)

Antenna	D (dB)	Z_A (Ω)	η_r (%)	Z_L (Ω)	Cross-Section (dBsw) (NEC4)
Straight-wire dipole	2.14	$72.2 + j0.99$	99.8	72.2	-6.74
Short dipole	1.93	$49.12 - j1.38$	99.8	50	-7.24
Folded spherical helix	1.75	$46.62 - j0.22$	98.2	50	-7.93

matched load, they will deliver exactly the same power to the load. From this perspective, there is no advantage to using one antenna relative to the other. From a practical perspective, other issues become significant, and there are some advantages associated with using a small loop rather than a small dipole.

Often, particularly in HF applications, the small loop is oriented vertically so as to receive vertical polarization. In this orientation, the loop will exhibit a bi-directional pattern providing some angular discrimination in azimuth that is unavailable with the vertically polarized small dipole. In general the small dipole will have higher radiation efficiency, but it will exhibit lower realized gain due to its higher mismatch loss (these antennas are typically used in 50- Ω systems). The small dipole exhibits a very high capacitive reactance (thousands of ohms), while the loop exhibits relatively low values of inductive reactance (tens or hundreds of ohms). In addition to its lower mismatch loss, the loop's low inductive reactance is easier to tune than the dipole's high capacitive reactance. With a loop, a relatively low loss adjustable capacitor can facilitate the tuning of the loop over a wide range of frequencies. With the dipole, a relatively lossy adjustable inductor would be needed to tune the high capacitive reactance.

In addition to using the equivalent circuit of Figure 6-16 to determine the captured power, you might consider using the circuit to determine the power scattered by the antenna (or scattering cross-section) using $P_s = \frac{1}{2} |I_L|^2 R_A$. While the Thevenin equivalent circuit can always be used to determine the power delivered to the load, it cannot always be used to determine the power scattered by the antenna (power dissipated in the source). In some instances, a Norton equivalent circuit is necessary to correctly determine scattered power. For many antennas, neither circuit can be used to determine the scattered power. While details of these concepts are beyond the scope of this chapter, it is important to consider the scattering cross-section of the small antenna relative to its much larger counterparts.

The results in Table 6-1 demonstrated that the power captured by the matched, efficient small antenna nearly equals the power captured by the much larger antenna, differing primarily as a function of the difference in antenna directivity. Similar behavior is also exhibited with the scattering cross-section of the small antenna. In general the scattering cross-section of the matched, efficient small antenna will nearly equal the scattering cross-section of the larger antenna, differing as a function of variations in directivity and residual structural scattering. A comparison of the scattering cross-section of the three antennas described in Table 6-1 is presented in Table 6-3.

6.7 THE EFFECTIVE VOLUME

Wheeler was the first to describe the fundamental relations between the performance properties of the small antenna and its occupied volume. Wheeler's discussion on the effective volume of the small antenna from the previous edition of the *Handbook* is included here in its entirety.

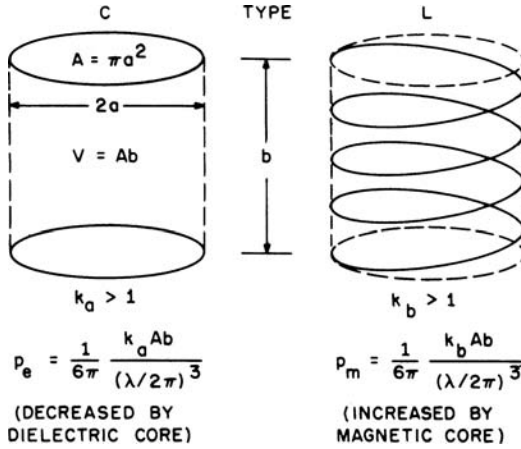


FIGURE 6-18 The radiation power factor in terms of volume

For any shape of small antenna of either kind (*C* or *L*), the radiation power factor at one frequency is proportional to the volume. Moreover, it is nearly equal for the two kinds if they occupy nearly equal volume.* This statement needs explanation, because their configuration differs in accordance with the properties of different materials.

Figure 6-18 shows examples of the two kinds configured to occupy the space within cylinders of equal dimensions. From familiar formulas for *C*, *L*, and radiation resistance, they have equal values of radiation power factor except for two factors (k_a , k_b), which are somewhat greater than unity. Either of these factors multiplies the volume ($V = Ab$) to give the effective volume as here defined (V').

In Figure 6-18 the effective volume is compared with the radian cube $(\lambda/2\pi)^3$. It is more logical to compare it with the volume of the radiansphere:

$$V_s = \frac{4\pi}{3} \left(\frac{\lambda}{2\pi} \right)^3 \quad (6-29)$$

Within this sphere, the stored energy or reactive power is predominant. Outside this sphere, the radiated power is predominant.

In terms of the effective volume ($V' = k_a Ab$ or $k_b Ab$ in Figure 6-18), the radiation power factor becomes

$$p = \frac{2}{9} \frac{V'}{V_s} \quad (6-30)$$

(The coefficient 2/9 reflects some properties of the near field of either antenna.) The effective volume may be stated as a sphere of radius a' :

$$V' = \frac{4\pi}{3} (a')^3 \quad p = \frac{2}{9} \left(\frac{2\pi a'}{\lambda} \right)^3 \quad a' = \frac{\lambda}{2\pi} \left(\frac{9}{2} p \right)^{\frac{1}{3}} \quad (6-31)$$

* This statement applies to the capacitive and inductive antennas described by Wheeler shortly. It does not apply to the lossless or lossy straight-wire dipole and circular-loop that occupy the same volume.

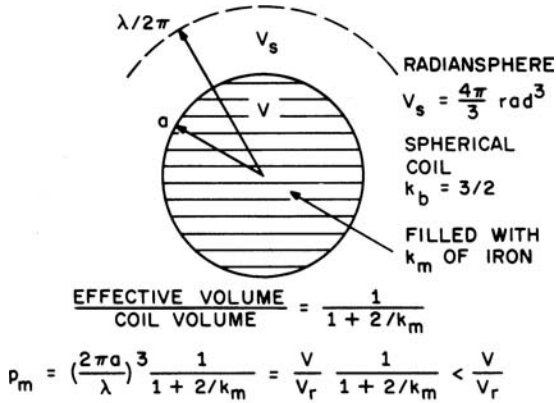


FIGURE 6-19 The spherical coil with a magnetic core

Note that a certain shape of self-resonant coil radiates equally from both C and L , so the total radiation is double that of either one. This is known as the *helix radiator of circular polarization* in the normal mode.

One theoretical case of a small antenna has the greatest radiation power factor obtainable within a spherical volume. Figure 6-19 shows such an antenna and its relation to the radiansphere (V_s). It is a spherical coil with a perfect magnetic core. The effective volume of an empty spherical coil has a shape factor $3/2$. Filling with a perfect magnetic core ($k_m = \infty$) multiplies the effective volume by three.

$$p_m = \frac{2}{9} \frac{(3)(3/2)V}{V_s} = \frac{V}{V_s} = \left(\frac{2\pi a}{\lambda}\right)^3 \quad (6-32)$$

The antenna is indicated by the shaded sphere a .

This idealized case depicts the physical meaning of the radiation power factor that cannot be exceeded. Outside the sphere occupied by the antenna, there is stored energy or reactive power that conceptually fills the radiansphere, but there is none inside the antenna sphere. The reactive power density, which is dominant in the radiation within the radiansphere, is related to the real power density, which is dominant in the radiation outside.

In a rigorous description of the electromagnetic field from a small dipole of either kind, the radiation of power in the far field is accompanied by stored energy that is located mostly in the near field (within the radiansphere). The small spherical inductor in Figure 6-19 is conceptually filled with perfect magnetic material, so there is no stored energy inside the sphere. This removes the avoidable stored energy, leaving only the unavoidable amount outside the inductor but mostly inside the radiansphere. This unavoidable stored energy is what imposes a fundamental limitation on the obtainable radiation power factor.

For any actual antenna, the effective volume and its spherical radius are

$$V' = \frac{9}{2} p V_s \quad a' = \frac{\lambda}{2\pi} \left(\frac{9}{2} p\right)^{\frac{1}{3}} \quad (6-33)$$

This volume includes any image in an adjoining ground plane regarded as integral with the antenna. It is convenient to show any antenna configuration with its sphere of effective volume drawn to scale, as a rating of its radiation power factor.

6.8 THE RADIATION SHIELD (THE WHEELER CAP)

Wheeler described an ingenious yet simple technique to measure the radiation efficiency of the small antenna. His discussion on the Wheeler Cap from the previous edition of the *Handbook* is included here in its entirety. Additional details on how to determine the small antenna's radiation efficiency using measured impedance with and without the Wheeler cap are included.

It is difficult to measure the small radiation power factor, and it is especially difficult to separate it from the loss power factor of the antenna and its tuning reactor. The radiation shield was devised to separate the two.

Figure 6-20 shows the concept of the radiation shield. Its purpose is to preserve the near field in the radiansphere while avoiding radiation farther out. Ideally it is a spherical box (of high conductivity) whose radius is $\lambda/2\pi$. Its dissipation is negligible as compared with free-space radiation.

The power factor of the tuned antenna is measured with and without the shield, to give the ratio

$$\frac{\text{Loss PF}}{\text{Loss PF} + \text{Radiation PF}} \quad (6-34)$$

From this ratio, the radiation power factor can be evaluated as a fraction of the measured total. The radiation efficiency of the antenna is given by

$$\eta_r = \frac{\text{Radiation PF}}{\text{Loss PF} + \text{Radiation PF}} \quad (6-35)$$

The radiation shield is not critical as to size or shape. A circular or square cylinder is usually convenient. It should be large enough to avoid much disturbance of the field near the antenna and small enough to avoid cavity resonance, especially in any mode excited by the antenna. If open-ended, it should be long enough to attenuate radiation outside. A small shift of antenna resonance frequency is tolerable.

With the Wheeler Cap, the impedance measurement of the tuned or self-resonant antenna in free-space provides the value of R_A . The measurement of the antenna with the

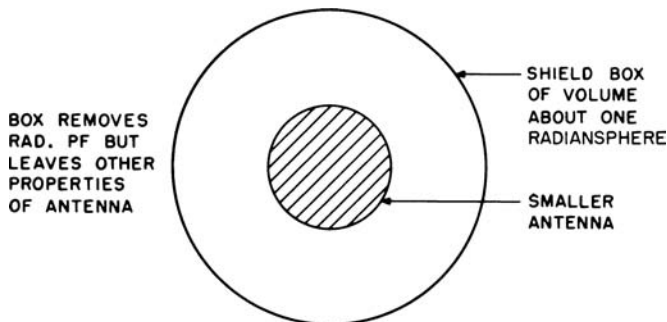


FIGURE 6-20 The radiation shield for use in measuring the radiation power factor

Wheeler Cap provides the value of R_l .^{*} If the loss resistance of the small antenna occurs in series with the radiation resistance (where $R_A > R_l$), the radiation efficiency of the antenna is found using

$$\eta_r = 1 - \frac{R_l}{R_A} \quad (6-36)$$

Equation 6-36 is valid for use with electrically small dipole antennas and electrically small loops [operating well below their first natural antiresonance ($ka \ll 0.5$)], where the loss resistance is in series with the radiation resistance.

If the loss resistance of the small antenna occurs in parallel with the radiation resistance (where $R_l > R_A$), the radiation efficiency of the antenna can be found using

$$\eta_r = 1 - \frac{R_A}{R_l} \quad (6-37)$$

Equation 6-37 is valid for use with electrically small loops (or other small antennas) operating near their antiresonant frequency where the radiation and loss resistances are in parallel. It is important when measuring (or simulating) with the Wheeler Cap that the measurements of both R_A and R_l be taken as close to their respective antiresonant frequencies as possible, where $X = 0$ in each case.

6.9 TYPICAL SMALL ANTENNAS

In this section, fundamental design approaches are described for several types of electrically small antennas. As noted previously, most if not all small antennas are derivatives of the straight-wire dipole and/or circular loop. The significant design challenge with a small antenna is impedance matching it to some characteristic impedance with wide operating bandwidth and high radiation efficiency. It was shown that small antennas can be designed to exhibit high radiation efficiency as a function of allowable volume and conductor diameter. Since the operating bandwidth of the small antenna is fundamentally limited by its overall volume, the small antenna design will reach a point where additional bandwidth cannot be realized. With all designs, there is always a trade-off that can be made between bandwidth and efficiency (gain). In general the widest bandwidth is achieved when the antenna is designed to occupy all of the available volume.

For these reasons this section primarily focuses on techniques that are used to match the small antenna to some desired characteristic impedance. A characteristic impedance of 50Ω is chosen. If the small antenna is to be connected to a device with a different characteristic impedance, 50Ω may not be the optimum design objective. The discussion that follows focuses on the basic monopole antenna since small dipole-like antennas typically have very high capacitive reactance and require more efficient tuning techniques. A monopole is chosen over the dipole since many practical antennas operate with some ground plane structure. The concepts described here apply equally to both monopoles

^{*} While the Wheeler Cap measurement of radiation efficiency can be accurate, some limitations must be considered. First, the cap must have sufficient radius so as not to disturb the near fields of the antenna. Second, losses within the cap are included in the value of R_l , which may reduce the accuracy of the measurement.

and dipoles. All of the numerical results presented in this section are obtained using EZNEC Pro.

Consider a straight-wire monopole antenna mounted over an infinite ground plane* having an overall height $h = 10$ cm and a conductor diameter $d = 1$ mm. The monopole antenna exhibits its first natural resonance at a frequency of 708.4 MHz having a total resistance of approximately 36Ω . At the frequency where $ka = 0.5$ (238.56 MHz), the monopole has an impedance $2.65 - j478.1 \Omega$ and a radiation efficiency of 98.4 percent. The approximate Q of the monopole at 238.56 MHz is 200. The height of the monopole is 0.0795λ at this frequency.

With decreasing frequency, the fixed height monopole exhibits decreasing radiation resistance, approaching half the value of Eq. 6-7 as $ka \rightarrow 0.5$ and lower. For a monopole of height h , Eq. 6-7 is rewritten as

$$R_r^m \approx 40\pi^2 \left(\frac{h}{\lambda} \right)^2 \quad (6-38)$$

To efficiently operate the small monopole in a 50- Ω system, its capacitive reactance must be tuned to 0 and its input resistance must be matched to 50 Ω . At low frequencies, where the monopole's height is much less than the operating wavelength, its input impedance has the form of $Z = R + jX$, where the reactance X is primarily a function of the conductor's self-inductance (L_s) and the capacitance between the conductor and the ground plane (C_G). These inductance and capacitance values are translated to the feedpoint, and the monopole's input reactance takes the form of

$$X = j \left(\omega L - \frac{1}{\omega C} \right) \quad (6-39)$$

where L and C are the values of L_s and C_G translated to the feedpoint, respectively. Given the short conductor length and geometry of the monopole, both L and C are small, and as $\omega \rightarrow 0$, the $-1/\omega C$ term dominates the behavior of the total reactance. To tune the total reactance to 0 at a given frequency ω_0 , the values of either or both of L and C can be increased.

The value of L can be increased by adding conductor length to the monopole in any geometry while maintaining the same overall height h . In adding conductor length, the objective is to increase the total series inductance at the feedpoint. Generally the most effective geometry for achieving the greatest inductance with a fixed wire length is the helical coil. Any arbitrary or defined (meander line, fractal, etc.) geometry can be used to tune the monopole's reactance to 0. Generally all other geometries require more wire length to achieve the same effective inductance that can be achieved with a helical coil of the same height and conductor diameter.

For this reason, the normal mode helix antenna, depicted in Figure 6-21a, is often used as a small antenna. The effect of the helical coil is the tuning of the monopole's capacitive reactance to 0. While the total resistance typically increases relative to that of the straight-wire monopole, it does not change substantially with the addition of conductor length to

* If the small antenna is mounted at the center of a ground plane with dimensions that are reasonably larger than the antenna, its impedance will be similar to that exhibited on the infinite ground plane. If the small antenna is mounted at the corner or edge of the ground plane, its impedance will be notably different.

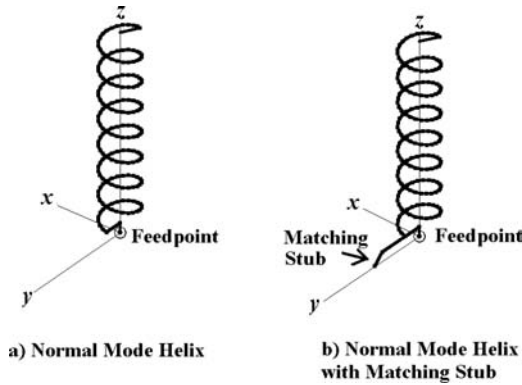


FIGURE 6-21 Depictions of (a) the unmatched normal mode helix and (b) the normal mode helix matched using a shunt matching stub

the antenna.* The normal mode helix is self-resonant at a frequency of 238.64 MHz having a resonant resistance of $6.2\ \Omega$. The radiation efficiency is 89.1 percent. The overall height and conductor diameter are the same as that of the monopole antenna. The total conductor length is 48.61 cm, which is approximately 0.387λ , slightly longer than the approximate 0.25λ length of a self-resonant straight-wire monopole. Generally more total wire length is needed to achieve resonance in a compressed geometry antenna.

Impedance matching of the normal mode helix can be accomplished using a parallel matching inductor at the feedpoint as discussed in previous sections. A simple technique to implement a parallel inductor at or near the feedpoint of the antenna is to connect a shunt stub between the antenna structure and the ground plane, as illustrated in Figure 6-21b. Typically only a short length of wire is required to achieve the impedance match. Minor frequency tuning can be realized by moving the connection point along the length of the antenna conductor.

The unmatched and matched impedances of the normal mode helix are presented in Figure 6-22 for a frequency range of 230 to 240 MHz. If the stub were placed exactly at the base of the feedpoint, the antenna would have been matched at a frequency lower than 237 MHz. Moving the location of the stub connection point along the antenna structure adjusts the match frequency. It is important to note that any small antenna exhibiting impedance properties similar to those shown in Figure 6-22a can be impedance matched using a shunt stub.

In this instance, the normal mode helix is matched at 237 MHz with a VSWR of 1.03 on $50\ \Omega$. The matched normal mode helix has a realized efficiency of approximately 98 percent. The Q of the matched normal mode helix is 53.2, substantially less than the Q of the straight-wire monopole. The lower bound on Q at this frequency is 10.37.

The high capacitive reactance of the monopole antenna can also be tuned to 0 by increasing the capacitance to ground. This is easily accomplished by adding a capacitive top-hat¹⁷

* This statement assumes that the antenna has a reasonable conductor diameter and that the geometry of the antenna is constrained to a reasonably narrow cylindrical diameter so that the capacitance between the antenna and the ground plane is not significantly affected. Additionally it is assumed that when winding the conductor beginning at the feedpoint, the value of height above the ground plane increases as you move along the conductor (the total antenna height remains h , independent of how much conductor is added). Geometries where the conductor is permitted to move in any direction (away from and toward the ground plane) can achieve higher values of resistance.

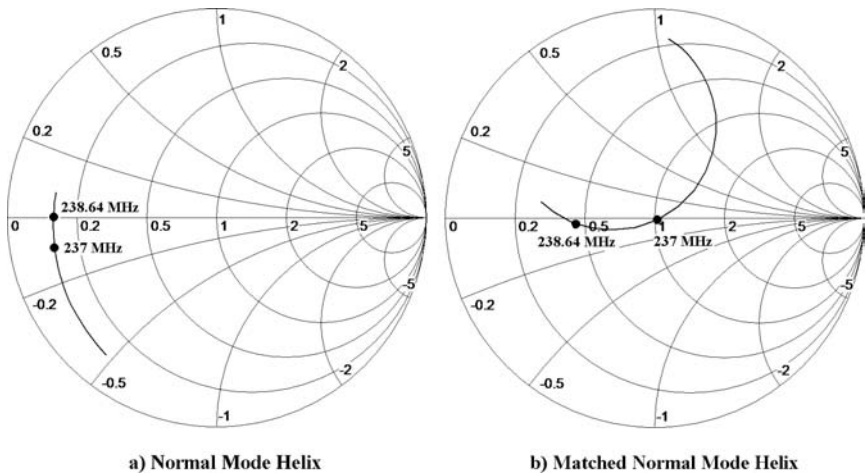


FIGURE 6-22 Impedance of (a) the unmatched normal mode helix and (b) the normal mode helix matched using a shunt matching stub

to the monopole, as depicted in Figure 6-23. Figure 6-23 illustrates a capacitive top-hat implemented using a wire grid structure rather than a solid disk so that direct comparison of the NEC4 simulations can be made to the normal mode helix.

To tune the reactance of the straight-wire monopole to 0 at a frequency near 238.56 MHz, a top-hat diameter of 10.62 cm is required. The capacitive top-hat antenna is self-resonant at a frequency of 238.44 MHz with a resistance of $9.63 \, \Omega$ and a radiation efficiency of 99 percent. Both values are higher than those of the normal mode helix antenna. The Q of the antenna is 16.47, substantially lower than that of the normal mode helix and straight-wire monopole. This antenna can be impedance matched in the same manner as the normal mode helix.

In general the capacitive top-hat design is more efficient and offers lower Q than the inductively loaded designs that have increased wire length compressed into a fixed volume. Note that the value of a (the spherical radius of the antenna's occupied volume) is larger with this antenna due to the increased cylindrical diameter occupied by the top-hat. The value of a for the capacitive top-hat antenna is 11.32 cm compared with the 10 cm for the normal mode helix. The lower bound on Q for the top-hat antenna is 7.44.

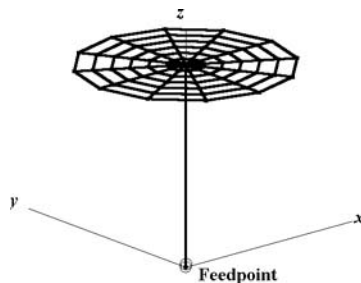


FIGURE 6-23 The capacitive top-hat antenna

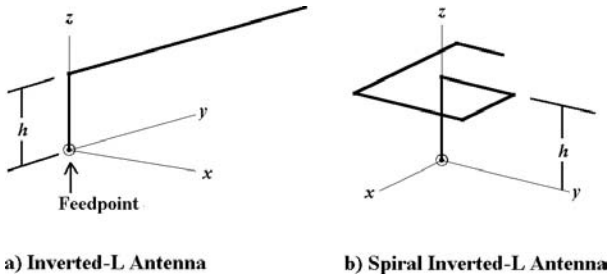


FIGURE 6-24 Depictions of (a) the inverted-L antenna and (b) the spiral inverted-L antenna

In the preceding paragraphs, the high capacitive reactance of a short straight-wire monopole was tuned to 0 using increased wire length or a capacitive top-hat. Each of these design approaches can be matched to $50\ \Omega$ using a parallel (shunt) matching stub. Now consider a straight-wire monopole that is naturally resonant near 238.56 MHz. This monopole has a height $h = 30.07$ cm and a conductor diameter $d = 1$ mm. The resonant resistance is approximately $36\ \Omega$. The design objective now is to reduce the height of the straight-wire monopole and to maintain high radiation efficiency and a reasonable match to $50\ \Omega$.

To reduce the height of the monopole, the conductor is folded over parallel to the ground plane as illustrated in Figure 6-24a. As the total height $h \rightarrow 0$, the monopole becomes the well-known inverted-L antenna.¹⁸ A performance summary of the antenna as a function of decreasing height is presented in Table 6-4. The total conductor length is held constant at 30.07 cm as h decreases. With decreasing values of h , there is a slight increase in resonant frequency. This is compensated for by adding an appropriate amount of length to the horizontal section of wire. More importantly, with decreasing height there is always a decrease in radiation resistance and an increase in Q . These variations are directly related to the change in $(h/\lambda)^2$.⁶

If the horizontal dimension (cylindrical diameter) of the inverted-L is too large, the conductor can be wound into a spiral as illustrated in Figure 6-24b. Any shift in resonant frequency can be corrected by adjusting the total conductor length. The total resistance of the

TABLE 6-4 Performance Properties of Reduced Height Monopole Antenna as a Function of Decreasing Height

Height (cm)	Resonant frequency (MHz)	Resonant resistance (Ω)	Q
30.07	238.6	36.1	7.6
25	243.7	34.3	8
20	246.9	29	9.4
15	249.1	20.9	12.9
10	250.1	11.7	22.5
5	250	3.7	65.8

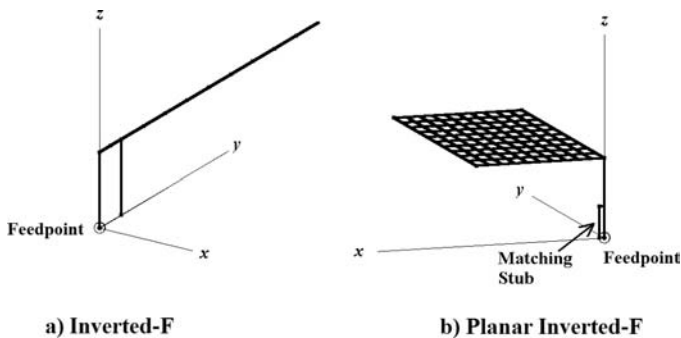


FIGURE 6-25 Depictions of (a) the inverted- F antenna and (b) the planar inverted- F antenna

spiral configuration does not shift significantly. The spiral version of the inverted-L antenna depicted in Figure 6-24b has a total conductor length of 35.7 cm and a conductor diameter of 1 mm. The resonant frequency and resistance are 238.54 MHz and $2.8\ \Omega$, respectively.

Impedance matching the inverted-L is accomplished using the shunt stub technique that was used with the normal mode helix antenna. Placing a shunt stub near the feedpoint of the inverted-L results in the well-known inverted- F antenna¹⁸ depicted in Figure 6-25a. The conductor length and stub position are adjusted to match the antenna at approximately 238.3 MHz with a VSWR of 1.3 on $50\ \Omega$. The radiation efficiency of the inverted- F is 93.1 percent. The Q of the antenna is 52.4, substantially lower than that of a straight-wire monopole of the same height.

Often, in an attempt to improve the bandwidth of a low profile inverted- F antenna, its volume is increased by converting the upper wire conductor into a plate, as illustrated in Figure 6-25b. This configuration of an inverted- F antenna is referred to as a *planar inverted- F antenna* (PIFA). PIFAs take many forms and are commonly used in a variety of wireless communication applications and devices.¹⁹ The PIFA illustrated in Figure 6-25b has a total height of 4.37 cm and top plate dimensions of 7.3 cm by 14.6 cm. It is matched at 238.5 MHz with a VSWR of 1.24 on $50\ \Omega$ and a radiation efficiency of 94.8 percent. The Q of the antenna is 47.8, slightly lower than that of the inverted- F .

It has been shown that small antennas can be designed to be reasonably efficient and matched to practical impedance values, provided there is sufficient available volume. In wireless devices, available volume is often limited, and it can be challenging to design efficient matched antennas. The design of a small antenna can be more challenging when required to operate over multiple operating bands. In many of these applications, acceptable performance is achieved by relaxing the VSWR requirements. At the end of the antenna design process, the single limiting performance characteristic for the small antenna is the maximum achievable operating bandwidth, which is fundamentally established by available volume.

REFERENCES

1. H. A. Wheeler, "Small Antennas," *IEEE Trans. Antennas Propagat.*, vol. 23 (July 1975): 462–469.
2. H. A. Wheeler, "The Radiansphere Around a Small Antenna," *Proc. of the IRE*, vol. 47 (August 1959): 1325–1331.

3. H. A. Wheeler, "Fundamental Limitations of Small Antennas," *Proc. of the IRE*, vol. 35 (December 1947): 1479–1484.
4. H. A. Wheeler, "A Helical Antenna for Circular Polarization," *Proc. of the IRE*, vol. 35 (December 1947): 1484–1488.
5. R. B. Adler, L. J. Chu, and R. M. Fano, *Electromagnetic Energy Transmission and Radiation* (New York: John Wiley & Sons, 1960).
6. S. R. Best, "On the Resonant Properties of the Koch Fractal and Other Bent Wire Monopole Antennas," *IEEE AWPL*, vol. 1, no. 3 (2002): 74–76.
7. S. R. Best, "On the Performance Properties of the Koch Fractal and Other Bent Wire Monopole Antennas," *IEEE Trans. Antennas Propagat.*, vol. 51, no. 6 (June 2003): 1292–1300.
8. R. Lewallen, EZNEC/4 Antenna Modeling Software, <http://www.eznec.com>.
9. C. A. Balanis, *Antenna Theory—Analysis and Design* (New York: John Wiley & Sons, 1982).
10. S. R. Best and A. D. Yaghjian, "The Lower Bound on Q for Lossy Electric and Magnetic Dipole Antennas," *IEEE AWPL*, vol. 3 (2004): 314–316.
11. S. R. Best, "The Foster Reactance Theorem and Quality Factor for Antennas," *IEEE AWPL*, vol. 3 (2004): 306–309.
12. L. J. Chu, "Physical Limitations of Omni-Directional Antennas," *Journal of Applied Physics*, vol. 10 (December 1948): 1163–1175.
13. A. D. Yaghjian and S. R. Best, "Impedance, Bandwidth and Q of Antennas," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 4 (April 2005): 1298–1324.
14. J. S. McLean, "A Re-Examination of the Fundamental Limits on the Radiation Q of Electrically Small Antennas," *IEEE Trans. Antennas Propagat.*, vol. 44 (May 1996): 672–676.
15. J. D. Kraus, "The Helical Antenna," *Proc. of the IRE*, vol. 37 (March 1949): 263–272.
16. D. M. Kerns, *Plane-Wave Scattering-Matrix Theory of Antennas and Antenna-Antenna Interactions* (Washington, DC: U.S. Government Printing Office, 1981).
17. E. W. Seeley, "An Experimental Study of the Disk Loaded Folded Monopole," *Proc. of the IRE*, vol. 44 (January 1956): 27–28.
18. R. King, C. W. Harrison, and D. H. Denton, Jr., "Transmission-Line Missile Antennas," *IRE Trans. Antennas Propagat.*, vol. 8 (January 1960): 88–90.
19. K.-L. Wong, *Planar Antennas for Wireless Communications* (Hoboken, NJ: Wiley-Interscience, 2003).

Chapter 7

Microstrip Antennas

David R. Jackson
University of Houston

CONTENTS

7.1 INTRODUCTION	7-2
7.2 BASIC PRINCIPLES OF OPERATION	7-4
7.3 CAD MODEL FOR THE INPUT IMPEDANCE	7-5
7.4 RADIATION PATTERNS	7-8
7.5 CAD FORMULAS FOR RECTANGULAR PATCH	7-11
7.6 RESULTS FOR RECTANGULAR PATCH	7-17
7.7 CAD FORMULAS FOR CIRCULAR PATCH	7-19
7.8 CIRCULAR POLARIZATION	7-22
7.9 MICROSTRIP ANTENNAS WITH IMPROVED PERFORMANCE	7-24

7.1 INTRODUCTION

Microstrip antennas (often called *patch* antennas) are widely used in the microwave frequency region because of their simplicity and compatibility with printed-circuit technology, making them easy to manufacture either as stand-alone elements or as elements of arrays. In its simplest form a microstrip antenna consists of a patch of metal, usually rectangular or circular (though other shapes are sometimes used) on top of a grounded substrate, as shown in Figure 7-1. In this chapter the basic principles of operation are discussed, and CAD formulas are given for the microstrip antenna. The CAD formulas are fairly accurate for thin substrates and illustrate the basic principles. For thin substrates the CAD formulas may even be accurate enough for final design purposes. For thicker substrates these formulas can still be used for initial design work, with full-wave simulation tools used to complete the final design.

History

The origin of microstrip antennas apparently dates back to 1953, when Deschamps proposed the use of microstrip feed lines to feed an array of printed antenna elements.^{1,2} The printed antenna elements introduced there were not microstrip patches, but flared planar horns. The microstrip patch antenna was first introduced by Munson in a symposium paper in 1972,³ which was followed by a journal paper in 1974.⁴ These papers discussed both the wraparound microstrip antenna and the rectangular patch. Shortly after Munson's symposium paper, Howell also discussed rectangular patch antennas in another symposium paper⁵ in which he credits Munson with the basic idea by referencing a private communication. In a later journal paper, Howell introduced the circular patch as well as the circularly polarized patch antenna.⁶ Soon after the introduction of the microstrip antenna, papers appeared describing methods of analysis for these antennas, including the transmission-line model,⁷ the cavity model,⁸ and the spectral-domain method.⁹ A good review of the early history of microstrip antennas is provided in the article by Carver and Mink.¹⁰ A discussion of microstrip antennas may be found in a variety of books devoted to this type of antenna¹¹⁻²² as well as in more general antenna books and handbooks.²³⁻²⁷

Feed Methods

Various methods may be used to feed the microstrip antenna, as shown in Figure 7-2 for the rectangular patch. The coaxial probe feed shown in Figure 7-2a is one of the most common feeds for a stand-alone element. The inset feed in Figure 7-2b is common for array applications. The proximity-coupled feed in Figure 7-2c requires multilayer fabrication,

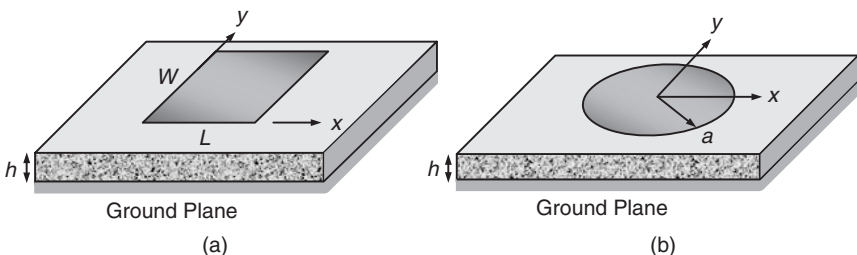


FIGURE 7-1 (a) Rectangular microstrip patch antenna and (b) circular microstrip patch antenna

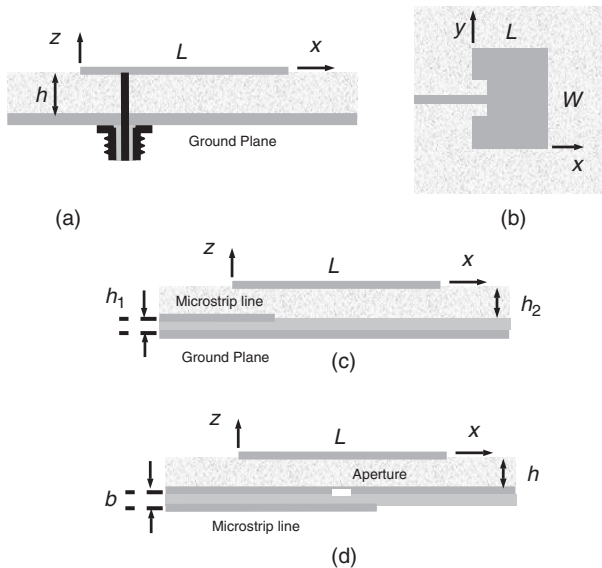


FIGURE 7-2 Feeding methods for a microstrip antenna: (a) coaxial feed, (b) inset feed, (c) proximity-coupled feed, and (d) aperture-coupled feed

but reduces spurious radiation from the feed line. The aperture-coupled feed shown in Figure 7-2d has the advantage of eliminating feed-line radiation (at the expense of some back radiation from the aperture) and also allows for relatively thick substrates, since probe reactance is not an issue.

Advantages and Disadvantages

Microstrip antennas usually have the important advantage of being low profile, and if the substrate is thin enough, they may also be *conformable*, meaning that the substrate can be bent to fit a curved surface, making the antenna very unobtrusive. Because the lateral size of a microstrip antenna on a substrate board is typically on the order of a half wavelength in the dielectric, size considerations usually dictate that these antennas are used in the UHF frequency band or higher, up through millimeter-wave frequencies, with microwave frequency applications being the most common. The main disadvantages of microstrip antennas include potentially lower radiation efficiency compared with other antennas (although this depends significantly on the substrate permittivity and thickness) and small bandwidth. These issues are discussed further next and in Section 7.9.

Radiation Efficiency and Bandwidth

Radiation efficiency depends largely on the substrate permittivity and thickness. A substrate that has a higher permittivity or that is thicker will suffer from increased surface-wave excitation, which will lower the efficiency. (Using a foam substrate is a simple way to eliminate surface-wave excitation. Removing the substrate outside of the patch cavity will also eliminate surface-wave excitation.) On the other hand, if the substrate is too

thin, the efficiency will be low due to conductive and dielectric losses. Assuming a typical Teflon substrate ($\epsilon_r = 2.2$ with a loss tangent of 0.001) and copper for the patch and ground plane with a conductivity of 3.0×10^7 S/m, the radiation efficiency is maximum for a substrate thickness of about $0.02\lambda_0$, reaching about 90 percent. When using a foam substrate, the efficiency continuously increases with increasing substrate thickness, approaching 100 percent for thicker substrates. More information on improving efficiency may be found in Section 7.9.

The bandwidth increases with the substrate thickness and inversely with the substrate permittivity, so bandwidth is made larger by using thicker low-permittivity substrates (e.g., a thicker foam substrate) at the expense of increased lateral size and vertical thickness. For a coaxially fed microstrip antenna, the substrate thickness is limited by the inductance of the feeding coaxial probe, which increases directly with increasing substrate thickness. For a typical substrate material such as Teflon and a 50- Ω coaxial feed cable, the probe reactance will become sufficiently large when the substrate thickness is about $0.018\lambda_0$ to render the antenna nonresonant unless a matching element is used. This limits the achievable bandwidth of a simple coaxially fed microstrip antenna. For a Teflon substrate and copper conductors, the maximum bandwidth (SWR < 2) of a rectangular microstrip antenna having a typical width/length (W/L) ratio of 1.5 will be about 2.5 percent, reached when the substrate thickness is about $0.025\lambda_0$. For a typical substrate thickness of $0.01\lambda_0$, the bandwidth is about 1.5 percent.

Many specialized techniques have been developed to increase the bandwidth of a microstrip antenna. These include either using thick foam substrates along with aperture-coupled feeds to avoid the probe reactance limitation, or using capacitive elements to compensate for the probe inductance. Even further increases may be achieved by using configurations that exhibit dual or multiple resonances, including stacked resonators or antennas surrounded by parasitically coupled elements. Antennas that have special geometries may also be used to greatly increase the bandwidth. By using these techniques, bandwidths exceeding 100 percent have been achieved. Further information about bandwidth improvement may be found in Section 7.9. Broadband printed antennas are the subject of Chapter 16, where you will find much more information about broadbanding techniques.

7.2 BASIC PRINCIPLES OF OPERATION

In this section, the basic operating principles of the microstrip antenna are reviewed. The mechanism of radiation is discussed first, to understand how the microstrip antenna radiates. The modes of operation of the rectangular and circular microstrip antennas are then reviewed. This provides the foundation for the calculation of the far-field radiation patterns of the rectangular and circular microstrip antennas, the results of which are presented in Section 7.4.

Radiation Mechanism

At first glance it might seem surprising that a microstrip antenna can operate very well at all, since it consists of a horizontal electric surface current (corresponding to the patch current) suspended (via the substrate) a short distance above a ground plane. Basic image theory predicts that such a current will not radiate very well. However, the microstrip patch and the ground plane together form a resonant cavity (filled with the substrate material). The cavity is lossy, due not only to the material (conductor and dielectric) loss, but also to the (desirable) radiation into space. Neglecting material loss, the quality factor Q of the antenna is inversely proportional to the substrate thickness h , for a given substrate material, assuming the substrate

is thin. Hence the bandwidth is proportional to h . The field level inside the patch cavity at resonance from an impressed current source inside the cavity (e.g., a fixed probe current) is proportional to Q . This means that the surface current on the patch (which is mainly on the lower surface of the patch) is inversely proportional to h . This increase in the amplitude of the surface current at resonance as the substrate gets thinner exactly balances the image effect, which causes the radiation level to be reduced by a factor proportional to h (relative to the patch current radiating without the ground plane). From another point of view, the voltage at the edges of the patch (the electric field times h) for a resonant patch remains approximately independent of h as the substrate gets thinner. Hence, both the electric and magnetic current models (discussed later) predict that the radiation from the patch remains approximately independent of h as the substrate thickness decreases.

Therefore, without material losses, the patch remains a good radiator even for very small substrate thicknesses, and it is possible to obtain a good impedance match even for a very thin substrate. In the lossless case the lower limit on the substrate thickness would only be determined by the bandwidth one is willing to accept. In actuality, the Q is limited by the material losses, so for sufficiently thin substrates it becomes difficult to obtain a good impedance match (in this region the radiation efficiency will also be poor). However, even for substrates as thin as $0.005\lambda_0$ a good match may be obtained with a reasonable efficiency of around 65 percent for a typical Teflon substrate and copper conductors.

Modes of Operation

For the rectangular patch, the TM_{mn} mode has a normalized electric field that is given by

$$E_z^{mn}(x, y) = \cos\left(\frac{m\pi x}{L}\right) \cos\left(\frac{n\pi y}{W}\right) \quad (7-1)$$

The usual mode of operation for a broadside pattern is the TM_{10} mode, which has no y variation and has a length L that is approximately one-half wavelength in the dielectric. In this mode the patch essentially acts as a wide microstrip line of width W that forms a transmission-line resonator of length L . The width W is usually larger than the length L in order to increase the bandwidth (see Eq. 7-35). A ratio $W/L = 1.5$ is typical.

For the circular patch, the TM_{np} mode has a normalized electric field that is given by

$$E_z(\rho, \phi) = \cos(n\phi) \frac{J_n\left(\frac{x'_{np}\rho}{a}\right)}{J_n(x'_{np})} \quad (7-2)$$

where x'_{np} is the p^{th} root of the Bessel function $J_n'(x)$. The usual mode of operation is the TM_{11} mode with $x'_{11} = 1.8418$. This mode has the lowest resonance frequency and has a broadside pattern.

7.3 CAD MODEL FOR THE INPUT IMPEDANCE

A simple circuit model for a probe-fed patch antenna operating near resonance is shown in Figure 7-3. The patch cavity resonator is modeled as a parallel RLC circuit (tank circuit), and the inductance due to the coaxial feed probe is modeled as the series inductor L_p . A justification for this model comes from an eigenfunction analysis of the input impedance.⁸

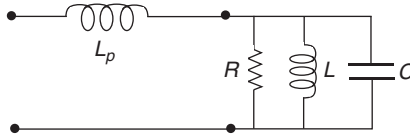


FIGURE 7-3 CAD circuit model for the probed microstrip patch antenna

From simple circuit theory, the input impedance of the patch is then given by

$$Z_{\text{in}} = j\omega L_p + \frac{R}{1 + jQ\left(f_R - \frac{1}{f_R}\right)} \quad (7-3)$$

where the frequency ratio is defined as $f_R = f/f_0$, with f_0 being the resonance frequency of the patch cavity (the resonance frequency of the RLC circuit). This is not the same as the impedance resonance frequency of the patch (the frequency for which the input reactance is zero), denoted as f_r , due to the presence of the probe inductance. The term R represents the input resistance of the patch at the cavity resonance frequency f_0 ($f_R = 1$), at which the input resistance is a maximum. CAD formulas for L_p , f_0 , Q , and R are given in Section 7.5 for a rectangular patch and in Section 7.7 for a circular patch. At the impedance resonance frequency f_r the input resistance will be slightly lower than the maximum value R according to the approximate formula

$$R_{\text{in}} = \frac{R}{1 + \left(\frac{X_p}{R}\right)^2} \quad (7-4)$$

where $X_p = \omega_0 L_p$ is the probe reactance. The probe reactance shifts the impedance resonance up from the cavity resonance by an amount $\Delta f = f_r - f_0$ given by the approximate formula

$$\frac{\Delta f}{f_0} = (BW) \left(\frac{1}{\sqrt{2}} \right) \left(\frac{X_p}{R} \right) \quad (7-5)$$

where

$$BW = \frac{1}{\sqrt{2}Q} \quad (7-6)$$

is the bandwidth of the antenna ($\text{SWR} < 2$ definition) and Q is the total quality factor. The input impedance of the tank circuit (the second term on the right-hand side of Eq. 7-3) along with its real and imaginary parts may be written in a normalized form as

$$\bar{Z}_{\text{RLC}} = \frac{1}{1 + jx} \quad \bar{R}_{\text{RLC}} = \frac{1}{1 + x^2} \quad \bar{X}_{\text{RLC}} = \frac{-x}{1 + x^2} \quad (7-7)$$

where

$$x = Q \left(f_R - \frac{1}{f_R} \right) \approx 2Q(f_R - 1) \quad (7-8)$$

is a normalized frequency term, and the bars over the impedance symbols denote that they have been normalized by dividing the impedances by R . The equations in Eq. 7-7 predict that the input resistance reaches a maximum of $R_{in} = R$ at the frequency f_0 ($x = 0$), while the magnitude of the reactance is maximum at $x = \pm 1$, for which $X_{in} = \mp R/2$. At f_0 the input impedance of the patch is simply $Z_{in} = R + jX_p$.

Using the CAD circuit model, the input impedance is plotted versus frequency for a typical rectangular patch in Figure 7-4, using the CAD formulas for the parameters of the circuit model given in the next section. Also shown is the result from a full-wave moment-method simulation using Ansoft Designer. This patch is designed to have an input resistance of $R = 50 \Omega$ at $f_r = 1.575$ GHz. The substrate has a relative permittivity of $\epsilon_r = 2.2$, a loss tangent of 0.001, and a thickness of $h = 0.1524$ cm (60 mils). The patch has an aspect ratio of $W/L = 1.5$ so that $L = 6.255$ cm and $W = 9.383$ cm. The ground plane and patch each have a conductivity of 3.0×10^7 S/m. The feed is an SMA coax with a probe radius $a = 0.635$ mm that is located at $x_0 = 1.85$ cm and $y_0 = W/2$. The CAD and full-wave results agree well except for a slight frequency shift. Note that the impedance resonance ($X_{in} = 0$) is shifted to a slightly higher frequency than the cavity resonance (where R is maximum), as expected, and the input reactance of the patch at f_0 is about 10Ω due to the probe reactance. The difference between the maximum and minimum values of the reactance curve is about the same as the value of R , as expected from Eq. 7-7. Taking the average of the maximum and minimum reactances provides an easy way to determine the probe reactance.

For this typical substrate thickness ($h/\lambda_0 = 0.008$), the simple CAD circuit model, together with the CAD formulas in the next section, works fairly well. For thicker substrates the circuit model will lose accuracy. The model is typically useful in the range $h/\lambda_0 < 0.03$. More sophisticated input impedance models include the transmission-line model⁷ and the cavity model⁸ (which may be implemented using either the eigenfunction-expansion method or the mode-matching method). However, these models are based on the same thin-substrate approximation as are the circuit model and the CAD formulas given in the next section, so the improvement in accuracy is questionable. (One advantage of the cavity model, however, is that it accounts for the probe reactance more

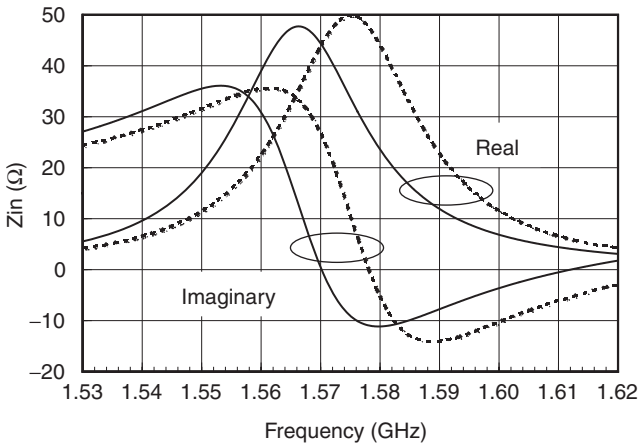


FIGURE 7-4 Input impedance calculated by the CAD model and by a full-wave moment-method simulation. The results from the full-wave simulation are shown with solid lines, while those from the CAD model are shown with dashed lines.

accurately than does the simple CAD formula for X_p given in the next section, which is independent of the feed position.) For thicker substrates a full-wave simulator is recommended for maximum accuracy.

7.4 RADIATION PATTERNS

The radiation patterns of a patch may be calculated using either an electric-current model or a magnetic-current model. These models are usually derived assuming that either the electric current on the patch or the electric field at the boundary of the patch corresponds to that of the dominant patch mode for a patch with ideal (magnetic-wall) boundaries. Both models then yield the same result for the far-field pattern when applied to a patch operating at the resonance frequency of the dominant mode of the ideal cavity.²⁸ The patterns may be calculated directly for two important cases. The first case assumes that the ground plane and substrate are both infinite. In the second case the ground plane is infinite, while the substrate is truncated at the edges of the patch. In the magnetic-current model the analysis for the latter case is the same as for the former case, except that the substrate material in the radiation model is now air. The radiation patterns are given next for the rectangular and circular patches, using the magnetic-current model.

Rectangular Patch

For the rectangular patch shown in Figure 7-1, the dominant TM_{10} mode has an electric field of the form

$$E_z(x, y) = \cos\left(\frac{\pi x}{L}\right) \quad (7-9)$$

The electric surface current corresponding to this cavity field is

$$\underline{J}_s(x, y) = \hat{x} \left(\frac{-1}{j\omega\mu} \right) \left(\frac{-\pi}{L} \right) \sin\left(\frac{\pi x}{L}\right) \quad (7-10)$$

and this is the current that produces the far-field radiation according to the electric-current model. In the magnetic-current model, the equivalence principle²⁹ is used to replace the patch with a magnetic surface current at the perimeter of the patch. The magnetic current has the form

$$\underline{M}_s = \begin{cases} -\hat{y} & x = L \\ -\hat{y} & x = 0 \\ -\hat{x} \cos\left(\frac{\pi x}{L}\right) & y = W \\ \hat{x} \cos\left(\frac{\pi x}{L}\right) & y = 0 \end{cases} \quad (7-11)$$

In the magnetic-current model the edges of the patch at $x = 0, L$ are referred to as the *radiating* edges since the magnetic currents on these edges are uniform and in phase. Those at $y = 0, W$ are referred to as the *nonradiating* edges since the magnetic currents on these edges reverse direction about the center of these edges and are also 180° out of phase on the two edges, so that these edges produce little radiation. The principal planes are denoted as the E -plane ($\phi = 0^\circ$) and the H -plane ($\phi = 90^\circ$). In the principal planes the nonradiating edges do not contribute when using the magnetic-current model.

As mentioned, the far-field pattern may be calculated assuming that the substrate is infinite or truncated at the edges of the patch. For thin substrates the truncation of the substrate does not have a significant effect on the pattern except near the horizon (θ approaching 90°) in the E -plane. For an infinite substrate the pattern will tuck in and go to zero at the horizon, while for a truncated substrate the pattern will remain nonzero down to the horizon in the E -plane.

A simple way to calculate the far-field pattern is by means of reciprocity.²⁹ The far field $E_i(r, \theta, \phi)$ ($i = \theta$ or ϕ) is calculated by placing a unit-amplitude testing dipole in the far field in the direction of interest (θ or ϕ). Since the testing dipole is in the far field, this calculation reduces to that of a plane wave impinging on the grounded substrate.³⁰ Such a plane-wave reflection problem has a simple closed-form solution. For the infinite substrate case, the far-field radiation pattern of the rectangular patch is given by

$$E_\theta^{\text{patch}}(r, \theta, \phi) = -2Wh \left(\frac{E_0}{\eta_0} \right) \cos \phi (1 - \Gamma^{\text{TM}}(\theta)) \cos \left(k_x \frac{L}{2} \right) \text{sinc} \left(k_y \frac{W}{2} \right) \text{tanc}(k_{z1}h) \quad (7-12)$$

$$E_\phi^{\text{patch}}(r, \theta, \phi) = 2Wh \left(\frac{E_0}{\eta_0} \right) (\cos \theta \sin \phi) (1 - \Gamma^{\text{TE}}(\theta)) \cos \left(k_x \frac{L}{2} \right) \text{sinc} \left(k_y \frac{W}{2} \right) \text{tanc}(k_{z1}h) \quad (7-13)$$

where

$$E_0 = \left(\frac{-j\omega\mu_0}{4\pi R} \right) e^{-jk_0 R} \quad (7-14)$$

and

$$\begin{aligned} k_x &= k_0 \sin \theta \cos \phi \\ k_y &= k_0 \sin \theta \sin \phi \end{aligned} \quad (7-15)$$

$$k_{z1} = k_0 N_1(\theta) \quad (7-16)$$

with

$$N_1(\theta) = \sqrt{\epsilon_r \mu_r - \sin^2 \theta} \quad (7-17)$$

and $\text{tanc}(x) = \tan(x)/x$, with η_0 being the intrinsic impedance of free space. In Eq. 7-14 the distance R is measured from the center of the patch. The terms Γ^{TM} and Γ^{TE} denote the TM and TE reflection coefficients as seen by a plane wave impinging from the air region on the interface between the air and substrate. For the infinite substrate model a simple calculation shows that

$$1 - \Gamma^{\text{TM}}(\theta) = \frac{2}{1 + j \left(\frac{N_1(\theta) \sec \theta}{\epsilon_r} \right) \tan(k_0 h N_1(\theta))} \quad (7-18)$$

and

$$1 - \Gamma^{\text{TE}}(\theta) = \frac{2}{1 + j \left(\frac{\mu_r \cos \theta}{N_1(\theta)} \right) \tan(k_0 h N_1(\theta))} \quad (7-19)$$

For a truncated substrate (truncated at the edges of the patch) and an infinite ground plane, the far field is given by the same Eqs. 7-12 and 7-13, except that now $\epsilon_r = 1$ and $\mu_r = 1$ are used in Eqs. 7-18 and 7-19. This also implies that $N_1(\theta) = \cos \theta$. The dimensions L and W used in Eqs. 7-12 and 7-13 are those of the original patch, however, corresponding to a resonant patch on the original substrate material (and not those of a resonant patch on an air substrate) even though air is used in Eqs. 7-18 and 7-19.

Figure 7-5 shows E - and H -plane patterns for a resonant rectangular patch on an infinite ground plane and substrate. The substrate permittivity is $\epsilon_r = 2.2$, the substrate thickness is $h/\lambda_0 = 0.02$, and the patch aspect ratio is $W/L = 1.5$. Note that the E -plane is broader than the H -plane, although in both planes the pattern tends to zero at the horizon.

Figure 7-6 shows patterns for the same antenna on a truncated substrate, where the substrate is truncated at the edges of the patch, while the ground plane remains infinite. The patterns are very similar to the infinite substrate case of Figure 7-5 except near the horizon. Near the horizon the H -plane patterns are similar, but the E -plane patterns are different. In the truncated case the E -plane pattern remains nonzero at the horizon, while the pattern tends to zero for the infinite substrate case. For the truncated substrate case the E -plane pattern will always remain nonzero at the horizon for any substrate material except air (in which case there is no difference between a truncated design and an infinite substrate design). For a patch on an air substrate the E -plane pattern will tend to zero at the horizon.

Circular Patch

The far-field patterns of a resonant circular patch are similar to those of the resonant rectangular patch on the same substrate material. The normalized field of the dominant TM_{11} mode is described by

$$E_z(\rho, \phi) = \cos(\phi) \frac{J_1\left(\frac{x'_{11}\rho}{a}\right)}{J_1(x'_{11})} \quad (7-20)$$

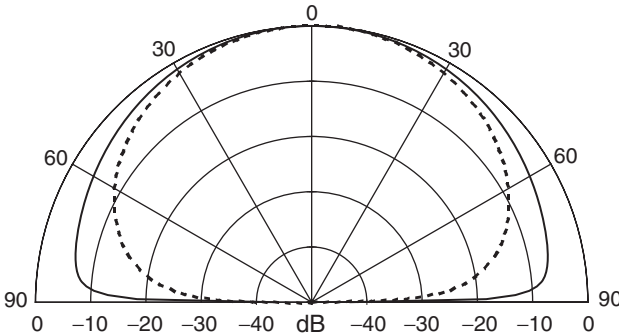


FIGURE 7-5 Far-field patterns of a rectangular patch on an infinite ground plane and substrate. The E -plane is shown with a solid line and the H -plane is shown with a dashed line.

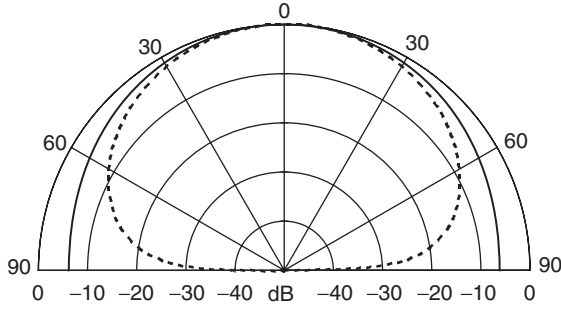


FIGURE 7-6 Far-field patterns of a rectangular patch on an infinite ground plane with a truncated substrate. The E -plane is shown with a solid line and the H -plane is shown with a dashed line.

This corresponds to a magnetic current at the edge of the patch that is

$$\underline{M}_s = \hat{\phi} \cos(\phi) \quad (7-21)$$

For this mode the far-field pattern is given by

$$E_\theta(r, \theta, \phi) = \left[2\pi \frac{E_0}{\eta_0} (ah) \right] \cos \phi \tan(k_{z1}h) J_1'(k_0 a \sin \theta) (1 - \Gamma^{\text{TM}}(\theta)) \quad (7-22)$$

$$E_\phi(r, \theta, \phi) = -2\pi \left[\frac{E_0}{\eta_0} (ah) \right] \sin \phi \tan(k_{z1}h) J_{\text{inc}}(k_0 a \sin \theta) (1 - \Gamma^{\text{TE}}(\theta)) \cos \theta \quad (7-23)$$

where E_0 is given by Eq. 7-14 (with R measured from the center of the patch) and

$$J_{\text{inc}}(x) = \frac{J_1(x)}{x} \quad (7-24)$$

The term $N_1(\theta)$ is defined in Eq. 7-17.

As with the rectangular patch, the patterns for a truncated patch are given by the same formulas as for the infinite substrate, in this case Eqs. 7-22 and 7-23, using $\epsilon_r = 1$ and $\mu_r = 1$. The dimension a is that of the original patch, however, and not that of a resonant patch on an air substrate.

Because the patterns of the circular patch are similar to those of the rectangular patch, the patterns are omitted here.

7.5 CAD FORMULAS FOR RECTANGULAR PATCH

This section presents CAD formulas for the rectangular patch shown in Figure 7-1. All of the formulas are independent of the feed except for the input resistance formula, which assumes a coaxial feed. All of these formulas assume that the patch is operating at the resonance of the TM_{10} mode, which is the usual mode of broadside operation. These formulas have been derived in Chapter 5 of Lee.¹⁷

Resonance Frequency

A fairly simple, yet reasonably accurate formula due to Hammerstad,³¹ is

$$f_0 = \frac{c}{2(L + 2\Delta L)\sqrt{\epsilon_r}} \quad (7-25)$$

where c is the speed of light in vacuum and the fringing extension added to the resonant (L) dimension is given by

$$\Delta L / h = \frac{0.412 (\epsilon_{\text{eff}} + 0.3) \left(\frac{W}{h} + 0.264 \right)}{(\epsilon_{\text{eff}} - 0.258) \left(\frac{W}{h} + 0.8 \right)} \quad (7-26)$$

with

$$\epsilon_{\text{eff}} = \frac{\epsilon_r + 1}{2} + \frac{(\epsilon_r - 1) \left(1 + 10 \frac{h}{W} \right)^{-1/2}}{2} \quad (7-27)$$

In some references ϵ_{eff} is used instead of ϵ_r in Eq. 7-25, but according to the discussion in Chapter 5 of Lee,¹⁷ using ϵ_r usually gives more accurate results.

Quality Factor

For a resonant type of antenna such as the microstrip antenna, it is common to express the physical parameters of interest in terms of the quality factor, or Q , of the antenna. The Q factor is defined as

$$Q = \omega_0 \left(\frac{U_s}{P_{\text{in}}} \right) \quad (7-28)$$

where $\omega_0 = 2\pi f_0$ is the resonance frequency in radian/s, U_s is the energy stored inside the patch cavity, and P_{in} is the average power going into the antenna, which is equal to the average power being radiated plus dissipated. A microstrip antenna has both dielectric and conductor losses, and possibly surface-wave loss as well. The surface-wave loss depends on the environment surrounding the patch. If there is a substrate that surrounds the patch and the surface-wave power launched by the antenna is gradually dissipated by an absorber, then the power launched into the surface wave by the patch is a loss. On the other hand, if the substrate is truncated so that there is no substrate beyond the perimeter of the patch, then there will be no surface-wave loss. If the substrate extends beyond the patch but then ends at some distance away, the propagating surface wave will diffract at the edge of the substrate and convert into a radiation field. In this case the surface wave power is not actually a loss, but the diffracted fields will typically result in pattern degradation.

The total quality factor is related to the component quality factors as

$$\frac{1}{Q} = \frac{1}{Q_{\text{sp}}} + \frac{1}{Q_{\text{sw}}} + \frac{1}{Q_d} + \frac{1}{Q_c} \quad (7-29)$$

where

$$Q_i = \omega_0 \left(\frac{U_s}{P_i} \right) \quad (7-30)$$

and Q_{sp} , Q_{sw} , Q_d , and Q_c denote the space-wave, surface-wave, dielectric, and conductor quality factors corresponding to the powers radiated into space, launched into the fundamental TM_0 surface-wave, dissipated by dielectric loss, and dissipated by conductor loss, respectively.

Dielectric Q If the substrate has a loss tangent denoted as $\tan \delta = \epsilon''/\epsilon'$, the dielectric Q factor is given simply by

$$Q_d = \frac{1}{\tan \delta} \quad (7-31)$$

Conductor Q If the ground plane and patch metal have surface resistances R_{sg} and R_{sp} , and R_s^{ave} denotes the average of the two, then the conductor Q factor is given by

$$Q_c = \mu_r \left(\frac{\eta_0}{2} \right) \left(\frac{k_0 h}{R_s^{ave}} \right) \quad (7-32)$$

The surface resistance is related to the conductivity of the metal σ and the skin depth δ as

$$R_s = \frac{1}{\omega \delta} \quad (7-33)$$

with

$$\delta = \sqrt{\frac{2}{\omega \mu_0 \sigma}} \quad (7-34)$$

Space-Wave Q The Q factor that accounts for the (desired) radiation into space is given as

$$Q_{sp} = \frac{3}{16} \left(\frac{\epsilon_r}{p_r c_1} \right) \left(\frac{L_e}{W_e} \right) \left(\frac{1}{h / \lambda_0} \right) \quad (7-35)$$

where $L_e = L + 2\Delta L$ and $W_e = W + 2\Delta W$ are the effective lengths and widths of the patch, accounting for fringing. The fringing length is given by Eq. 7-26, while the fringing width is given approximately by³²

$$\Delta W = h \left(\frac{\ln 4}{\pi} \right) \quad (7-36)$$

The terms p_r and c_1 are geometry terms given by (see Chapter 5 of Lee¹⁷)

$$\begin{aligned} p_r = & 1 + \frac{a_2}{10} (k_0 W_e)^2 + (a_2^2 + 2a_4) \left(\frac{3}{560} \right) (k_0 W_e)^4 \\ & + c_2 \left(\frac{1}{5} \right) (k_0 L_e)^2 + a_2 c_2 \left(\frac{1}{70} \right) (k_0 W_e)^2 (k_0 L_e)^2 \end{aligned} \quad (7-37)$$

$$c_1 = 1 - \frac{1}{n_1^2} + \frac{2}{n_1^4} \quad (7-38)$$

where $n_1 = \sqrt{\epsilon_r \mu_r}$ is the index of refraction of the substrate. The other constants are

$$a_2 = -0.16605$$

$$a_4 = 0.00761$$

$$c_2 = -0.0914153$$

Surface-Wave Q Assuming that the substrate is infinite (or that there is an absorber to absorb the surface wave), so that surface-wave power is a loss from the antenna radiation point of view, the surface-wave excitation represents a loss mechanism, and the Q for this loss is given by

$$Q_{sw} = Q_{sp} \left(\frac{e_r^{sw}}{1 - e_r^{sw}} \right) \quad (7-39)$$

where e_r^{sw} denotes the radiation efficiency of the patch when accounting for only surface-wave loss, and not dielectric or conductor loss. That is,

$$e_r^{sw} = \frac{P_{sp}}{P_{sp} + P_{sw}} \quad (7-40)$$

where P_{sp} is the power radiated into space and P_{sw} is the power launched into the surface wave. This efficiency is well approximated by that of a unit-amplitude infinitesimal horizontal electric dipole (*hed*) on the substrate. (Unit amplitude means $I l = 1$, with the current amplitude I expressed using peak phasor notation, not rms, and the length l of the small dipole measured in meters.) Therefore this efficiency is given by the approximate expression

$$e_r^{sw} = e_r^{hed} = \frac{P_{sp}^{hed}}{P_{sp}^{hed} + P_{sw}^{hed}} \quad (7-41)$$

A calculation that is accurate for thin substrates reveals that

$$P_{sp}^{hed} = \frac{1}{\lambda_0^2} (k_0 h)^2 (80 \pi^2 \mu_r^2 c_1) \quad (7-42)$$

and

$$P_{sw}^{hed} = \frac{1}{\lambda_0^2} (k_0 h)^3 \left[60 \pi^3 \mu_r^3 \left(1 - \frac{1}{n_1^2} \right)^3 \right] \quad (7-43)$$

which results in the expression

$$e_r^{\text{sw}} = e_r^{\text{hed}} = \frac{1}{1 + (k_0 h) \left(\frac{3\pi}{4} \right) \mu_r \frac{1}{c_1} \left(1 - \frac{1}{n_1^2} \right)^3} \quad (7-44)$$

For nonmagnetic substrates, a more accurate expression that has been found by Pozar³³ for the surface-wave power of the unit-amplitude dipole is

$$P_{\text{sw}}^{\text{hed}} = \frac{\eta_0 k_0^2}{4} \frac{\epsilon_r (x_0^2 - 1)^{\frac{3}{2}}}{\epsilon_r [1 + x_1] + (k_0 h) \sqrt{x_0^2 - 1} [1 + \epsilon_r^2 x_1]} \quad (7-45)$$

where

$$x_1 = \frac{x_0^2 - 1}{\epsilon_r - x_0^2}$$

and

$$x_0 = 1 + \frac{-\epsilon_r^2 + \alpha_0 \alpha_1 + \epsilon_r \sqrt{\epsilon_r^2 - 2\alpha_0 \alpha_1 + \alpha_0^2}}{(\epsilon_r^2 - \alpha_1^2)}$$

with

$$\alpha_0 = s \tan[(k_0 h) s]$$

$$\alpha_1 = -\frac{1}{s} \left[\tan[(k_0 h) s] + \frac{(k_0 h) s}{\cos^2[(k_0 h) s]} \right]$$

and

$$s = \sqrt{\epsilon_r - 1}$$

Using Eqs. 7-45 and 7-42 in Eq. 7-40 results in an expression that is more accurate than Eq. 7-44, especially for thicker substrates.

Bandwidth

The bandwidth of the patch may be defined from the frequency limits at which the standing-wave ratio (SWR) reaches a maximum threshold, assuming that the feeding transmission line

that connects to the patch is perfectly matched at the resonance frequency (i.e., $Z_0 = R$, if we neglect the effects of the probe inductance). The bandwidth is thus defined as

$$BW = \frac{f_2 - f_1}{f_r} \quad (7-46)$$

where f_r is the impedance resonance frequency of the patch, and f_1 and f_2 are the low and high frequencies on either side of the resonance frequency at which $SWR = S$, with S being a prescribed value. The bandwidth is then

$$BW = \frac{S-1}{\sqrt{S}Q} \quad (7-47)$$

For the commonly used definition $S = 2$, we have

$$BW = \frac{1}{\sqrt{2}Q} \quad (7-48)$$

Radiation Efficiency

The radiation efficiency is the ratio of power radiated into space to the total power input to the antenna. That is,

$$e_r = \frac{P_{sp}}{P_{in}} \quad (7-49)$$

where

$$P_{in} = P_{sp} + P_{sw} + P_d + P_c \quad (7-50)$$

In terms of the Q factors,

$$e_r = \frac{Q}{Q_{sp}} \quad (7-51)$$

Input Resistance

The resistance R in the circuit model of Figure 7-3 represents the input resistance of the patch at the cavity resonance frequency f_0 where the input resistance is a maximum. An approximate expression for R is

$$R = R_{edge} \cos^2 \left(\frac{\pi x_0^e}{L_e} \right) \quad (7-52)$$

where the input resistance when fed at the edge ($x_0 = 0$) is

$$R_{edge} = \eta_0 \mu_r \left(\frac{4}{\pi} \right) \left(\frac{L_e}{W_e} \right) \left(\frac{h}{\lambda_0} Q \right) \quad (7-53)$$

The effective feed location at, $x_0^e = x_0 + \Delta L$, $y_0^e = y_0 + \Delta W$ accounts for fringing.

Probe Inductance

A simple yet reasonably accurate formula for the probe inductance L_p in the circuit model of Figure 7-3 may be found by assuming that the coaxial probe feeds an infinite parallel-plate waveguide.³⁴ A calculation then reveals that the probe reactance is given by

$$X_p = \omega L_p = \eta_0 \mu_r \left(\frac{h}{\lambda_0} \right) \left[-\gamma + \ln \left(\frac{2}{\sqrt{\mu_r \epsilon_r} k_0 a} \right) \right] \quad (7-54)$$

where $\gamma = 0.57722$ is Euler's constant. Note that the probe inductance is not independent of frequency, due to the wavenumber term that appears inside the argument of the natural logarithm function. The probe reactance is directly proportional to the substrate thickness. The probe reactance also increases as the probe radius decreases, but the reactance is not a strong function of the probe radius since the variation is logarithmic. For a typical substrate with $\epsilon_r = 2.2$ and $h = 0.1524$ cm (60 mils) and a probe radius of $a = 0.635$ mm (SMA), the probe reactance X_p at 1.575 GHz is about 11 Ω .

Directivity

The directivity of the rectangular patch (with respect to an isotropic radiator) may be approximated in closed form for thin substrates as

$$D = \left(\frac{\eta_0}{40\pi} \right) \left(\frac{1}{p_r c_1} \right) \left[\frac{\text{tanc}^2(k_0 h n_1)}{1 + \left(\frac{\mu_r}{\epsilon_r} \right) \tan^2(k_0 h n_1)} \right] \quad (7-55)$$

where

$$\text{tanc}(x) = \tan(x)/x \quad (7-56)$$

For a moderate permittivity substrate such as $\epsilon_r = 2.2$, the directivity is about 6.1 (7.8 dB) when the substrate is thin. For a high permittivity substrate such as $\epsilon_r = 10.8$, the directivity is about 3.5 (5.4 dB) when the substrate is thin.

7.6 RESULTS FOR RECTANGULAR PATCH

Results are presented here to illustrate two of the basic physical properties of the rectangular patch that were discussed earlier: bandwidth and radiation efficiency. The substrate has a relative permittivity of either 2.2 (moderate permittivity) or 10.8 (high permittivity), a loss tangent of 0.001, and the patch and ground planes both have a conductivity of 3.0×10^7 S/m. The patch is resonant with an aspect ratio of $W/L = 1.5$. The frequency is 5.0 GHz.

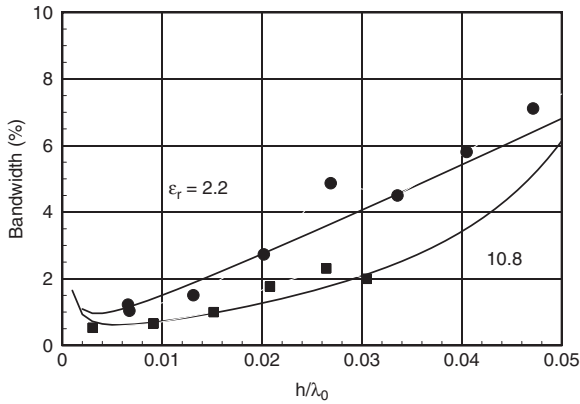


FIGURE 7-7 Bandwidth (percent) versus the normalized substrate thickness for a moderate permittivity substrate and a high-permittivity substrate. The solid curves are from the CAD formula, while the dots (moderate permittivity) and squares (high permittivity) are from measurements.

Figure 7-7 shows a plot of bandwidth versus normalized substrate thickness for the moderate and high permittivity substrates. The bandwidths calculated from the CAD formulas are compared with those from measurements. The CAD formula is seen to be fairly accurate over this range of substrate thickness. As predicted by the CAD formula, the bandwidth increases roughly proportional to the substrate thickness, and inversely proportional to the substrate permittivity.

Figure 7-8 shows a plot of radiation efficiency in percent versus normalized substrate thickness for the same two substrate materials. In Figure 7-8a the dielectric and conductor

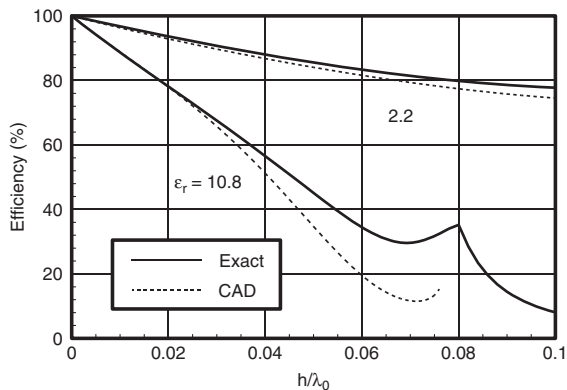


FIGURE 7-8a Radiation efficiency (percent) versus the normalized substrate thickness for a moderate-permittivity substrate and a high-permittivity substrate. The radiation efficiency accounts for surface-wave loss only. The dashed curves are from the CAD formula, while the solid curves are exact values from a spectral-domain calculation.

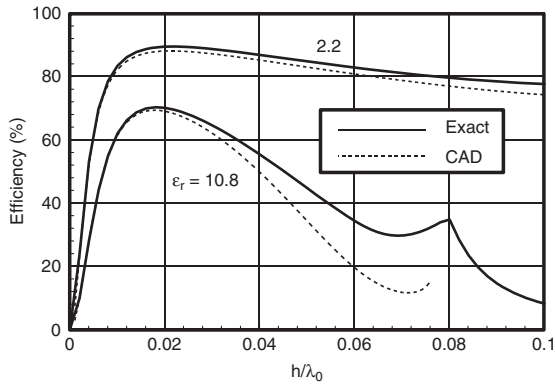


FIGURE 7-8b Radiation efficiency (percent) versus the normalized substrate thickness for a moderate-permittivity substrate and a high-permittivity substrate. The radiation efficiency accounts for all losses. The dashed curves are from the CAD formula, while the solid curves are exact values from a spectral-domain calculation.

losses are assumed to be zero, so radiation efficiency accounts only for surface-wave loss as in Eq. 7-40, using Eq. 7-45 in the calculation.

In Figure 7-8b the total radiation efficiency is plotted, accounting for all losses. In both cases the results from the CAD formulas are compared with the exact values, obtained from a spectral-domain calculation.³⁵ The CAD formula works well for thinner substrates, but loses accuracy for thicker substrates, especially for the higher permittivity.

When dielectric and conductor losses are neglected, the efficiency approaches 100 percent for thin substrates, since surface-wave loss becomes negligible. When dielectric and conductor losses are accounted for, the efficiency reaches a maximum for a certain substrate thickness, around 0.02 free-space wavelengths, and then decreases to zero as the substrate thickness tends to zero. This can be explained by examining how the different quality factors behave versus substrate thickness. The space-wave Q is inversely proportional to the electrical thickness of the substrate, as seen from Eq. 7-35. The surface-wave Q is inversely proportional to the square of the electrical thickness of the substrate, as seen from Eqs. 7-39 and 7-44. The dielectric Q is constant and does not vary with substrate thickness. The conductor Q is proportional to the substrate thickness, as seen from Eq. 7-32. Hence, when conductor and dielectric losses are eliminated, the space-wave Q becomes dominant (the smallest) as the substrate becomes thinner. However, when losses are considered, the conductor Q eventually becomes the dominant one as the substrate becomes thinner. This means that most of the input power to the antenna goes to conductor losses instead of into radiation into space, and the efficiency becomes small.

7.7 CAD FORMULAS FOR CIRCULAR PATCH

CAD formulas for the basic properties of the circular patch are given next. These CAD formulas are derived assuming that the substrate is electrically thin. Because the properties of the circular patch are similar to those of the rectangular patch, results are omitted.

Resonance Frequency

The resonance frequency of the TM_{11} mode is given by

$$f_0 = \frac{c}{2\pi a_e \sqrt{\epsilon_r}} x'_{11} \quad (7-57)$$

where c is the speed of light in vacuum and a_e is the effective radius of the patch, given by³⁶

$$a_e = a \sqrt{1 + \frac{2h}{\pi a \epsilon_r} \left[\ln \left(\frac{\pi a}{2h} \right) + 1.7726 \right]} \quad (7-58)$$

Quality Factor

The formulas for the dielectric and conductor quality factors Q_d and Q_c are the same as for the rectangular patch and are given by Eqs. 7-31 and 7-32. The space-wave quality factor is given by

$$Q_{sp} = \frac{2x_c \left(\frac{1}{k_0 h} \right) \epsilon_r}{I_c} \quad (7-59)$$

where

$$x_c \equiv \frac{1}{x'_{11}} (x'^2_{11} - 1) = 0.238695 \quad (7-60)$$

and

$$I_c \equiv \int_0^{\pi/2} C(\theta) d\theta \quad (7-61)$$

with

$$C(\theta) = \sin \theta \tanh^2(k_{z1} h) \left[\left| 1 - \Gamma^{TM}(\theta) \right|^2 J_1'^2(k_0 a \sin \theta) + \left| 1 - \Gamma^{TE}(\theta) \right|^2 \cos^2 \theta J_{inc}^2(k_0 a \sin \theta) \right] \quad (7-62)$$

The terms appearing in the preceding equation are the same as those defined in Eqs. 7-56, 7-16, 7-18, 7-19, and 7-24. An approximate closed-form expression for the space-wave quality factor may be obtained for a thin substrate by approximating the integral I_c in Eq. 7-61 as

$$I_c = \frac{4}{3} p_c \quad (7-63)$$

where

$$p_c = \sum_{k=0}^6 (k_0 a)^{2k} e_{2k} \quad (7-64)$$

with the coefficients defined as

$$\begin{aligned}
 e_0 &= 1 \\
 e_2 &= -0.400000 \\
 e_4 &= 0.0785710 \\
 e_6 &= -7.27509 \times 10^{-3} \\
 e_8 &= 3.81786 \times 10^{-4} \\
 e_{10} &= -1.09839 \times 10^{-5} \\
 e_{12} &= 1.47731 \times 10^{-7}
 \end{aligned} \tag{7-65}$$

The surface-wave quality factor is given by Eqs. 7-39 and 7-41, along with 7-42 and either 7-43 (which yields Eq. 7-44) or 7-45.

Bandwidth

The bandwidth of the circular patch is given by Eqs. 7-47 and 7-48, where the total quality factor is given by Eq. 7-29, using the appropriate formulas for the quality factors of the circular patch presented in the preceding section.

Radiation Efficiency

The radiation efficiency of the circular patch is given by Eq. 7-41 together with Eq. 7-42 and either Eq. 7-43 (which yields Eq. 7-44) or 7-45.

Input Resistance

The input resistance of the circular patch fed by a coaxial feed at the frequency of the cavity resonance f_0 is

$$R_{\text{in}} \approx R_{\text{edge}} \left[\frac{J_1^2 \left(\frac{x'_{11} \rho}{a} \right)}{J_1^2(x'_{11})} \right] \tag{7-66}$$

where the feed is located at a radius ρ_0 from the center. The edge resistance is given by

$$R_{\text{edge}} = \left[\frac{h^2}{2P_{\text{sp}}} \right] e_r \tag{7-67}$$

where

$$P_{sp} = \frac{\pi}{8\eta_0} (k_0 a)^2 h^2 \int_0^{\pi/2} \text{tanc}^2(k_{z1} h) \left[|1 - \Gamma^{\text{TM}}(\theta)|^2 J_1'^2(k_0 a \sin \theta) + |1 - \Gamma^{\text{TE}}(\theta)|^2 J_{inc}^2(k_0 a \sin \theta) \cos^2 \theta \right] \sin \theta d\theta \quad (7-68)$$

In Eq. 7-67, e_r is the radiation efficiency of the circular patch and P_{sp} is the space-wave power radiated by the magnetic current in Eq. 7-21. The space-wave power may be written as

$$P_{sp} = \frac{\pi}{8\eta_0} (k_0 a)^2 h^2 I_c \quad (7-69)$$

where the integral I_c is defined in Eq. 7-61 and approximated in Eq. 7-63.

Probe Inductance

Equation 7-54 is based on a parallel-plate waveguide model, which ignores the geometry of the patch. Therefore, this formula may be used equally well for the circular patch.

Directivity

An approximate formula for the broadside directivity of the circular patch is

$$D = \frac{3 \text{tanc}^2(k_1 h)}{p_c \left[1 + \left(\frac{\mu_r}{\epsilon_r} \right) \tan^2(k_1 h) \right]} \quad (7-70)$$

where p_c is defined in Eq. 7-64 and the tanc function is defined in Eq. 7-56.

7.8 CIRCULAR POLARIZATION

Microstrip antennas are often used as elements to radiate circular polarization (CP). Circular polarization may be achieved in two ways. The first is by means of a dual feed using a square or circular patch that is fed with two feeds spaced 90° apart and fed with a 90° phase difference. The 90° phase difference may be realized using a delay line or, for a broader bandwidth phasing, a branchline hybrid.³⁶ In most cases the bandwidth of the patch is small compared with the bandwidth of the phasing, so a simple 90° delay line is adequate.

A more compact way to realize CP is through a single feed, using a nearly-square patch or an elliptical patch. The nearly-square design has length L and width W (see Figure 7-1) that are nearly, but not exactly, equal. The feed is located along the diagonal. This method of realizing CP is simple. However, the drawbacks include having a lower CP bandwidth compared with the dual-feed arrangement, and the design may be somewhat more difficult to optimize properly. Simple design formulas exist, however, to aid in the design of the patch dimensions and the feed location. This is discussed next.

It is assumed that CP is desired at broadside at a frequency f_{cp} (the axial ratio will degrade as the angle θ increases). Denote f_{0x} and f_{0y} as the cavity resonance frequencies corresponding to the dimensions L and W , respectively. The first design equation is then

$$f_{cp} = \frac{1}{2}(f_{0x} + f_{0y}) \quad (7-71)$$

which stipulates that the frequency of operation should be halfway between the two cavity resonance frequencies. The second design equation says that the difference in the two cavity resonance frequencies should be related to the Q of the patch as

$$\frac{\Delta f}{f_{cp}} = \frac{1}{Q} \quad (7-72)$$

where $\Delta f = f_{0y} - f_{0x} > 0$ for left-handed circular polarization (LHCP) and $\Delta f = f_{0x} - f_{0y} > 0$ for right-handed circular polarization (RHCP). Hence, the resonance frequencies of the patch should be chosen as

$$f_{0x} = f_{cp} \left[1 \pm \frac{1}{2Q} \right] \quad (7-73)$$

$$f_{0y} = f_{cp} \left[1 \mp \frac{1}{2Q} \right] \quad (7-74)$$

where the top sign is for RHCP and the bottom sign is for LHCP. The quality factor Q is related to the impedance bandwidth BW of the patch when it is fed along a centerline and operated as a linearly polarized patch, through Eq. 7-6. The bandwidth of the linearly polarized patch may be determined by using the CAD formulas for the quality factor given previously, or by other methods including full-wave simulation and measurements.

The input resistance at the frequency f_{cp} is given by $R = R_x = R_y$, where R_x and R_y are the resonant input resistances of the patch when fed along the centerlines at $(x_0, W/2)$ or $(L/2, y_0)$ with $x_0 = y_0$, so that the (1,0) or (0,1) modes are resonant, respectively. Hence the input resistance of the diagonal feed at f_{cp} is simply the resonant input resistance if the patch were fed along either of the two centerlines, assuming that we keep the same value of x_0 or y_0 . For example, we can determine the value of x_0 for a square patch fed along the centerline in order to have a $50\text{-}\Omega$ input resistance at resonance. The input resistance of the nearly square patch fed along the diagonal will then be $50\text{ }\Omega$ at f_{cp} provided the distance of the feed from the corner of the patch measured along the diagonal is $x_0\sqrt{2}$. At the frequency of CP operation f_{cp} the input impedance will have a small imaginary part equal to the probe reactance.

Once the frequencies f_{0x} and f_{0y} have been determined, the corresponding dimensions of the patch L and W may be determined. For example, we can use Eq. 7-25, which becomes

$$f_{0x} = \frac{c}{2(L + 2\Delta L)\sqrt{\epsilon_r}} \quad (7-75)$$

$$f_{0y} = \frac{c}{2(W + 2\Delta W)\sqrt{\epsilon_r}} \quad (7-76)$$

Because the patch is nearly square, the fringing extensions ΔL and ΔW are approximately equal. The fringing extension ΔL may be calculated approximately using Eq. 7-26.

A bandwidth of the circular polarization may be defined in terms of the axial ratio of the CP at broadside. The axial-ratio bandwidth is defined as

$$BW^{AR} = \frac{f^+ - f^-}{f_{cp}} \quad (7-77)$$

where f^+ and f^- are the frequency limits on either side of f_{cp} for which the axial ratio is $\sqrt{2}$ (3 dB). It may be shown³⁷ that the axial-ratio bandwidth is related to the quality factor of the patch as

$$BW^{AR} = \frac{0.348}{Q} \quad (7-78)$$

Comparing Eq. 7-78 with Eq. 7-6, it is seen that the axial-ratio bandwidth is smaller than the impedance bandwidth of the corresponding linearly polarized patch by a factor of about 0.49. The impedance bandwidth of the nearly square CP patch is given by³⁷

$$BW^{CP} = \frac{\sqrt{2}}{Q} \quad (7-79)$$

Comparing Eq. 7-79 with Eq. 7-6, it is seen that the impedance bandwidth of the CP patch is larger than that of the same patch operated as a linearly polarized patch by a factor of 2.

7.9 MICROSTRIP ANTENNAS WITH IMPROVED PERFORMANCE

The improvement of the basic microstrip antenna properties is a topic of considerable interest. In particular, considering the traditional limitations of microstrip antennas, the improvements of bandwidth and radiation efficiency are particularly relevant, along with reducing surface-wave excitation and lateral radiation from the patch, which results in undesirable mutual coupling and edge diffraction from the edges of the surrounding ground plane. These issues are reviewed next.

Bandwidth Improvement

The bandwidth of a microstrip antenna may be improved in several ways. First, the use of a low permittivity substrate is beneficial since bandwidth is inversely proportional to the permittivity according to Eq. 7-35, at least when the patch is operating in a region where the efficiency is high and the space-wave Q is dominant (the smallest). Second, a thicker substrate may be used since the bandwidth is proportional to the substrate thickness in the region where the space-wave Q dominates, as Eq. 7-35 again shows. For a coaxial feed, the bandwidth is limited by the fact that the probe inductance increases with increasing substrate thickness, as seen from Eq. 7-54. As mentioned in the “Introduction,” this limits

the substrate thickness to a maximum value, beyond which the coaxially fed patch will not be resonant (i.e., the input impedance will remain inductive). According to Eq. 7-7 and the discussion immediately afterward, the input reactance of the patch cavity reaches a negative minimum of $X_{\text{RLC}} = -R/2$, where R is the value of the maximum input resistance, which occurs at the cavity resonance frequency f_0 . Hence if the probe reactance exceeds $R/2$, no resonance frequency f_r can be found for which the input impedance will be real.

Various methods may be used to overcome this limitation due to the probe reactance. A capacitive element may be inserted into the coaxial feed to compensate for the probe reactance. For example, one can use a capacitive annular slot on the patch that surrounds the contact point where the probe meets the patch.³⁸ Alternatively, other feed methods may be used that avoid the probe reactance problem. For example, the aperture-coupled feed allows for the patch cavity to be excited without introducing a probe reactance, and hence a thicker substrate is possible.

Other techniques to increase bandwidth include introducing multiple resonances into the structure. This may take the form of stacked patches, coplanar parasitic patches, or patches that have novel shapes such as the *U*-shaped patch.^{19,21} Using special feed networks or feeding techniques to compensate for the natural impedance variation of the patch is another method.

Chapter 16 of this *Handbook* is concerned with wideband printed antennas, and you will find more discussion on broadbanding techniques in that chapter.

Efficiency Improvement

Surface-wave excitation increases as the permittivity of the substrate increases and as the thickness of the substrate increases (unless the substrate is air). On the other hand, if the substrate is made too thin, the radiation efficiency suffers due to increased conductor and dielectric losses, since the conductor and dielectric Q factors will become dominant (smaller than the other ones) as the substrate becomes thin, as seen from Eqs. 7-31 and 7-32. One way to overcome this trade-off is to use a low-permittivity substrate material such as foam. The foam substrate may be made thick to minimize conductor and dielectric losses, without suffering from surface-wave excitation. This has the added advantage of increasing the bandwidth. If the application allows it, the substrate may also be removed from the area outside the patch, to avoid surface-wave excitation.

Other alternatives have been developed for minimizing surface-wave excitation. The ground plane may be patterned to form a periodic structure, or a periodic structure may be printed on the substrate surrounding the patch, to form an *electromagnetic band gap* or EBG structure.³⁹ Being a periodic structure, the EBG structure has passbands and stopbands, for which modal surface-wave propagation on the structure is allowed or prohibited. In the stopband region the surface wave on the structure cannot propagate. Care must be exercised to design the EBG structure so that the stopband is omnidirectional, meaning that the surface-wave propagation is prohibited in all directions.

Reduced Surface Wave Antennas

A type of microstrip antenna called the *reduced surface wave* (RSW) antenna has been proposed.⁴⁰ The design consists of a microstrip annular ring with inner radius a and outer radius b that is short-circuited to the ground at its inner radius, as shown in Figure 7-9. The short circuit may be realized by a circular ring of shorting vias (metal posts) placed at $\rho = a$.

The antenna is operated in the TM_{11} mode, for which the magnetic current at the outer boundary corresponding to the dominant patch mode has the form of Eq. 7-21.

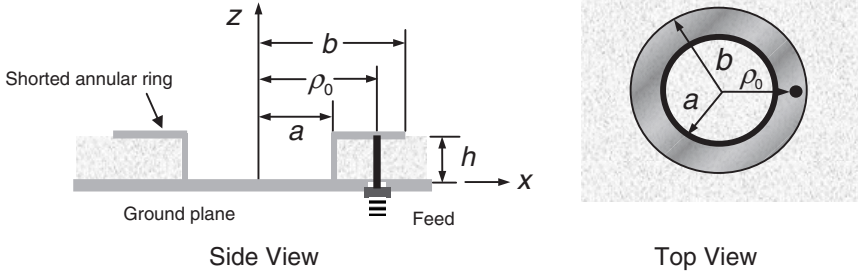


FIGURE 7-9 Reduced surface wave microstrip antenna

Analysis shows that the surface-wave field along the ground plane ($z = 0$) excited from such a ring of magnetic current will have the form

$$E_z^{\text{sw}}(\rho, \phi, 0) = A \cos \phi J_1'(\beta_{\text{TM}_0} b) H_1^{(2)}(\beta_{\text{TM}_0} \rho) \quad (7-80)$$

where A is an amplitude coefficient, β_{TM_0} is the wavenumber of the dominant TM_0 surface-wave mode of the grounded substrate, J_1' is the derivative of the Bessel function J_1 , and $H_1^{(2)}$ is the Hankel function of the second kind. The surface-wave field excited by this magnetic current will be zero provided that

$$\beta_{\text{TM}_0} b = x'_{11} = 1.8418 \quad (7-81)$$

To be more general, any value x'_{1p} may be used in Eq. 7-81, but using $p = 1$ gives the patch with the smallest outer radius. A normal circular patch with the critical radius b will not be resonant, as is easily seen by comparing Eq. 7-81 with Eq. 7-57, which is equivalent to Eq. 7-81 with k_1 being used instead of the wavenumber of the surface wave. A circular patch with this critical radius will be larger than resonant size. Therefore, to make the patch resonant, a short-circuit boundary is introduced at $\rho = a$ as shown in Figure 7-9. By properly selecting the inner radius a , the patch may be made resonant at the desired operating frequency for which Eq. 7-81 is satisfied. Assuming an approximate perfect magnetic wall at the outer radius, this results in the following design condition for choosing the inner radius:

$$\frac{J_1(k_1 a)}{Y_1(k_1 a)} = \frac{J_1' \left(\frac{k_1 x'_{11}}{\beta_{\text{TM}_0}} \right)}{Y_1' \left(\frac{k_1 x'_{11}}{\beta_{\text{TM}_0}} \right)} \quad (7-82)$$

where Y_1 is the Bessel function of the second kind. The location of the coaxial feed probe may be selected to obtain the desired input resistance at resonance.

If the substrate is relatively thin, the TM_0 surface wave has the property that the wavenumber is close to the wavenumber of free space. Hence we have the approximate condition that

$$k_0 b = x'_{11} = 1.8418 \quad (7-83)$$

Equation 7-83 is a condition for having zero lateral radiation from the patch, i.e., zero far-field radiation along the horizon. Hence the RSW antenna not only reduces the amount of surface-wave excitation, but it also reduces the lateral radiation that travels along the ground plane. For a thin substrate the lateral radiation is usually the more important of the two, unless the distance from the patch becomes large, but both contribute to diffraction from the edges of the ground plane. This undesirable edge diffraction results in interference with the primary pattern above the ground plane, and in nonzero radiation in the region behind the ground plane.

Equation 7-83 also implies that the diameter of the RSW antenna will be approximately $0.6\lambda_0$, independent of the substrate permittivity. That is, the RSW antenna cannot be miniaturized by using a higher substrate permittivity. This is one of the main drawbacks to the antenna. The antenna has roughly the same bandwidth as a conventional circular patch on the same substrate.

Figure 7-10a shows the calculated and measured E -plane patterns for a conventional circular patch of radius $a = 2.66$ cm operating at 2.0 GHz on a substrate with a relative permittivity of 2.94 and a thickness of 1.524 mm. The feed is at $\rho_0 = 0.61$ cm. The calculated pattern assumes an infinite ground plane and substrate, while the measurement was taken on a circular ground plane with a diameter of 1 m. The measured E -plane pattern has a significant rippling in the forward direction and has a significant amount of back radiation behind the ground plane, due to edge diffraction. This is because the surface wave and the lateral wave are both excited with a $\cos \phi$ variation, and hence the E -plane ($\phi = 0$) is the one that suffers the most from surface-wave excitation and lateral radiation. Also, the measured E -plane pattern does not agree very well with the calculated pattern, since the calculated pattern assumes an infinite ground plane and therefore neglects edge diffraction.

Figure 7-10b shows the calculated and measured E -plane patterns for an RSW antenna on the same substrate, operating at the same frequency. The dimensions are $a = 2.417$ cm and $b = 4.392$ cm, with the feed located at $\rho_0 = 2.95$ cm. It is seen that the E -plane pattern is much smoother in the forward region than for the conventional patch; also there is much less radiation in the back region behind the ground plane. Because there is less edge diffraction, the measured pattern agrees much better with the calculated pattern, which assumes an infinite ground plane and neglects edge diffraction.

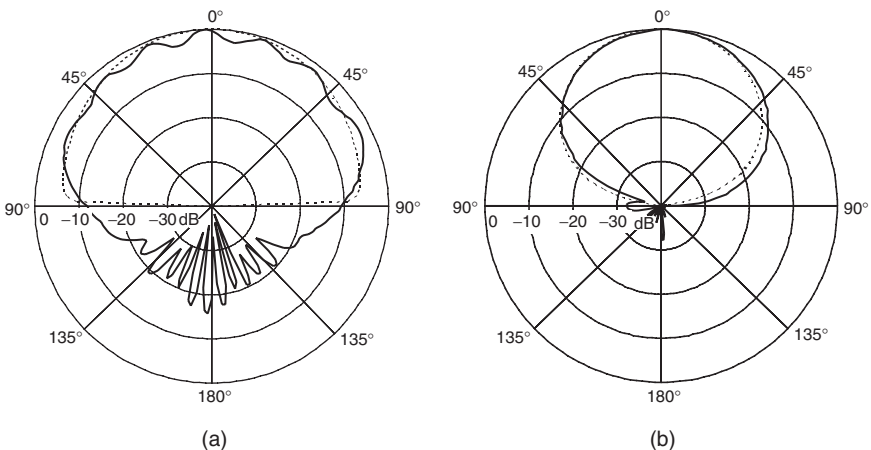


FIGURE 7-10 E -plane patterns for (a) conventional circular patch antenna (b) reduced surface wave antenna. The measured patterns are shown with solid lines, while the calculated patterns are shown with dashed lines.

As a consequence of the reduced surface-wave excitation and lateral radiation, there is also less mutual coupling between RSW antennas than between conventional microstrip antennas, at least for separations that are larger than about $0.75\lambda_0$ between centers.⁴¹ For array applications the large size of the RSW antenna limits the application of these antennas to arrays that can tolerate this larger element spacing.⁴²

Acknowledgment

The author would like to thank Varada Rajan Komanduri for performing the full-wave simulations and carefully proofreading the manuscript.

REFERENCES

1. G. Deschamps and W. Sichak, "Microstrip Microwave Antennas," *Proc. of Third Symp. on USAF Antenna Research and Development Program*, October 18–22, 1953.
2. J. T. Bernhard, P. E. Mayes, D. Schaubert, and R. J. Mailloux, "A Commemoration of Deschamps' and Sichak's 'Microstrip Microwave Antennas': 50 Years of Development, Divergence, and New Directions," *Proc. of the 2003 Antenna Applications Symp.* (September 2003): 189–230.
3. R. E. Munson, "Microstrip Phased Array Antennas," *Proc. of Twenty-Second Symp. on USAF Antenna Research and Development Program*, October 1972.
4. R. E. Munson, "Conformal Microstrip Antennas and Microstrip Phased Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-22, no. 1 (January 1974): 74–78.
5. J. Q. Howell, "Microstrip Antennas," *IEEE AP-S Intl. Symp. Digest* (1972): 177–180.
6. J. Q. Howell, "Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (January 1974): 90–93.
7. A. G. Derneryd, "Linearly Polarized Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 6 (November 1976): 846–850.
8. Y. T. Lo, D. Solomon, and W. F. Richards, "Theory and Experiment on Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-27, no. 3 (March 1979): 137–145.
9. M. D. Deshpande and M. C. Bailey, "Input Impedance of Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 4 (July 1982): 645–650.
10. K. R. Carver and J. W. Mink, "Microstrip Antenna Technology," *IEEE Trans. Antennas Propagat.*, vol. AP-29, no. 1 (January 1981): 25–38.
11. I. J. Bahl and P. Bhartia, *Microstrip Antennas* (Dedham, MA: Artech House, 1980).
12. J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna Theory and Design*, INSPEC/IEE (London and New York: Peter Peregrinus on behalf of the IEE, 1981).
13. J. R. James and P. S. Hall, *The Handbook of Microstrip Antennas* (two-volume set), INSPEC (London: Peter Peregrinus on behalf of the IEE, 1989).
14. P. Bhartia, *Millimeter-Wave Microstrip and Printed Circuit Antennas* (Boston: Artech House, 1991).
15. D. M. Pozar and D. H. Schaubert (eds.), *Microstrip Antennas: The Analysis and Design of Microstrip Antennas and Arrays* (New York: Wiley/IEEE Press, 1995).
16. Jean-Francois Zurcher and Fred E. Gardiol, *Broadband Patch Antennas* (Boston: Artech House, 1995).
17. K. F. Lee (ed.), *Advances in Microstrip and Printed Antennas* (New York: John Wiley & Sons, 1997).
18. R. Garg, P. Bhartia, I. J. Bahl, and A. Ittipiboon (eds.), *Microstrip Antenna Design Handbook* (Boston: Artech House, 2001).
19. G. Kumar and K. P. Ray, *Broadband Microstrip Antennas* (Boston: Artech House, 2002).

22. R. B. Waterhouse, *Microstrip Patch Antennas: A Designer's Guide* (Norwell, MA: Kluwer Academic Publishers, 2003).
21. K.-L. Wong, *Compact and Broadband Microstrip Antennas* (New York: John Wiley & Sons, 2003).
22. R. Bancroft, *Microstrip and Printed Antenna Design* (Atlanta: Noble Publishers, 2004).
23. J. D. Kraus, *Antennas*, 2nd Ed. (New York: McGraw-Hill, 1988).
24. W. F. Richards, "Microstrip Antennas," *Antenna Handbook: Theory, Applications, and Design*, Chap. 10, Y. T. Lo and S. W. Lee (eds.) (New York: Van Nostrand Reinhold, 1993).
25. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd Ed. (New York: John Wiley & Sons, 1998).
26. D. R. Jackson, J. T. Williams, and D. R. Wilton, "Antennas-II," *Handbook of Engineering Electromagnetics*, Chap. 9, Rajeev Bansal (ed.) (New York: Marcel Dekker, 2004).
27. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd Ed. (New York: John Wiley & Sons, 2005).
28. D. R. Jackson and J. T. Williams, "A Comparison of CAD Models for Radiation from Rectangular Microstrip Patches," *Intl. Journal of Microwave and Millimeter-Wave Computer Aided Design*, vol. 1, no. 2 (April 1991): 236–248.
29. R. F. Harrington, *Time Harmonic Electromagnetic Fields* (New York: McGraw-Hill, 1961).
30. D. R. Jackson and N. G. Alexopoulos, "Gain Enhancement Methods for Printed Circuit Antennas," *IEEE Trans. Antennas Propagat.*, vol. 33 (September 1985): 976–987.
31. E. O. Hammerstad, "Equations for Microstrip Circuit Design," *Proc. 5th European Microwave Conf.* (1975): 268–272.
32. H. A. Wheeler, "Transmission-Line Properties of Parallel Strips Separated by a Dielectric Sheet," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-13 (1965): 172–185.
33. D. M. Pozar, "Rigorous Closed-Form Expression for the Surface-Wave Loss of Printed Antennas," *Electron. Lett.*, vol. 26 (1990): 954–956.
34. H. Xu, D. R. Jackson, and J. T. Williams, "Comparison of Models for the Probe Inductance for a Parallel Plate Waveguide and a Microstrip Patch," *IEEE Trans. Antennas Propagat.*, vol. 53 (October 2005): 3229–3235.
35. D. M. Pozar, "Considerations for Millimeter-Wave Printed Antennas," *IEEE Trans. Antennas Propagat.*, vol. 31 (1983): 740–747.
36. D. M. Pozar, *Microwave Engineering*, 3rd Ed. (New York: John Wiley & Sons, 2005).
37. W. L. Langston and D. R. Jackson, "Impedance, Axial-Ratio, and Receive-Power Bandwidths of Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 52 (October 2004): 2769–2773.
38. P. S. Hall, "Probe Compensation in Thick Microstrip Patches," *Electron. Lett.*, vol. 11 (1987): 606–607.
39. H. Y. D. Yang, R. Kim, and D. R. Jackson, "Design Considerations for Modeless Integrated Circuit Substrates Using Planar Periodic Patches," *IEEE Trans. Microwave Theory Tech.*, vol. 48 (December 2000): 2233–2239.
40. D. R. Jackson, J. T. Williams, A. K. Bhattacharyya, R. Smith, S. J. Buchheit, and S. A. Long, "Microstrip Patch Designs That Do Not Excite Surface Waves," *IEEE Trans. Antennas Propagat.*, vol. 41 (August 1993): 1026–1037.
41. M. A. Khayat, J. T. Williams, D. R. Jackson, and S. A. Long, "Mutual Coupling Between Reduced Surface-Wave Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 48 (October 2000): 1581–1593.
42. R. L. Chen, D. R. Jackson, J. T. Williams, and S. A. Long, "Scan Impedance of RSW Microstrip Antennas in a Finite Array," *IEEE Trans. Antennas Propagat.*, vol. 53 (March 2005): 1098–1104.

Chapter 8

Slot Antennas

William F. Croswell

Harris Corporation

CONTENTS

8.1	INTRODUCTION.....	8-2
8.2	SLOTTED-WAVEGUIDE ANTENNAS	8-2
8.3	TAPERED AND FLARED SLOT ANTENNAS	8-4
8.4	CAVITY-BACKED RECTANGULAR SLOT ANTENNAS	8-5
8.5	WAVEGUIDE-FED SLOT ANTENNAS	8-8
8.6	SLOT ANTENNAS ON FINITE AND CURVED GROUND PLANES	8-10
8.7	SLOT ANTENNAS ON CYLINDERS, CONES, AND SPHERES.....	8-14

8.1 INTRODUCTION

This chapter deals with the radiation characteristics of slot antennas and includes the effects of finite and curved surfaces, dielectric coatings, cavity backing, and single slots fed by waveguides. The simplest example of such an antenna consists of a rectangular slot cut in an extended thin flat sheet of metal with the slot free to radiate on both sides of this sheet, as shown in Figure 8-1. The slot is excited by a voltage source such as a balanced parallel transmission line connected to the opposite edges of the slot or a coaxial transmission line connected to the opposite edges of the slot or a coaxial transmission line.

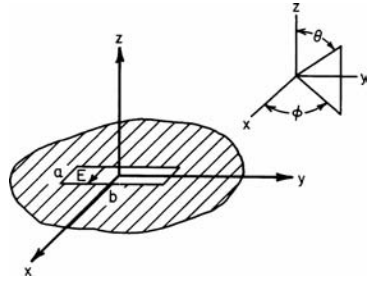


FIGURE 8-1 Rectangular slot

The electric field distribution in the slot can be obtained from the relationship between the slot and complementary wire antennas, as established by Booker.¹ It has been shown that the electric field distribution (magnetic current) in the slot is identical to the electric current distribution on the complementary wire. In the illustrated rectangular slot, the electric field is perpendicular to the long dimension, and its amplitude vanishes at the ends of the slot.

The electric field is everywhere normal to the surface of the slot antenna except in the region of the slot itself. The theoretical analysis of this configuration shows that the radiation of the currents in the sheet can be deduced directly from the distribution of the electric field in the slot. Consequently, the radiated field of an elementary magnetic moment within the slot boundaries should include the contribution of the electric current flowing on a metal surface.

A slot-antenna design will often require that the slot be cut in something other than an extended flat sheet surface. Whatever the surface, the electric field will be normal everywhere except in the region of the slot. The field due to the electric currents on this metal surface can be deduced from the exciting magnetic currents² in the slot, just as in the case of the flat metal sheet. This field can be combined with the exciting field so that the result is the total field due to a magnetic current on the given boundary surface. Thus the field of a thin rectangular slot cut in a circular cylinder differs from that of a slot cut in a flat metal sheet because the distribution of electric currents is different for the two cases.

In general, the slot antenna is not free to radiate on both sides of the surface on which it is cut because one side is either completely enclosed, e.g., the slotted cylinder antenna, or it is desired that the radiation on one side be minimized. In these cases, the influence of the enclosed cavity region on the excitation and impedance of the slot antenna is significant to the antenna design.

8.2 SLOTTED-WAVEGUIDE ANTENNAS

Slotted-waveguide antennas have significant applications in the areas of missile, spacecraft, and airborne radar. Broad-wall slotted-waveguide antennas have been studied extensively. Oliner,³ following research by Stevenson,⁴ has derived equivalent circuit representations for this type of antenna. These circuits allow accurate computation of the normalized conductance, susceptance, resistance, reactance, and resonant frequency of broad-wall slots. The types of slots and the equivalent networks are given in Chapter 9.

Dielectric-covered broad-wall slots also have been characterized thoroughly. Bailey⁵ has summarized significant findings. The general problem discussed by Bailey extends Oliner's work to include the effects on broad-wall slots radiating into a multilayer dielectric medium. The basic effect that occurs by adding a dielectric layer over slots in a waveguide is to

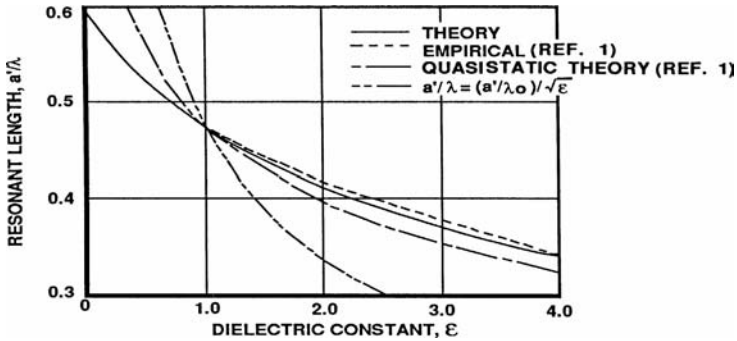


FIGURE 8-2 Resonant length of a dielectric-covered shunt slot in WR-90 (RG 52/U) waveguide ($b^1 = 0.0625$ in, $t = 0.050$ in, $a = 0.900$ in, $b = 0.400$ in) (after Bailey⁵ © IEEE 1970)

produce a downward shift in resonant frequency. For dielectric layers whose thickness is greater than 0.2λ , where $\lambda = \lambda_0 \sqrt{\epsilon}$, the approximate resonant shift is given in Figure 8-2, where λ is the wavelength in the dielectric, λ_0 is the wavelength in free space, and ϵ is the dielectric constant. Further analysis, and comparison with experimental data, is also available, as shown in Figure 8-3.

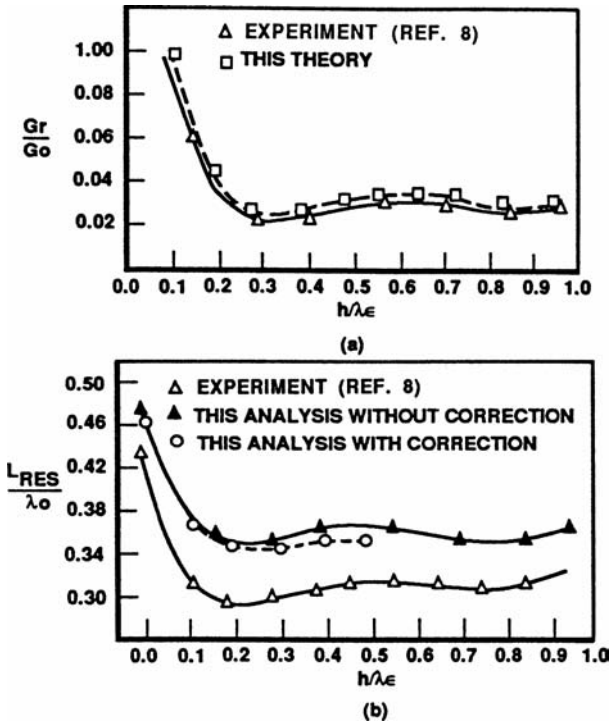


FIGURE 8-3 Resonant conductance (a) and resonant length (b) versus layer thickness, $\epsilon_r = 3.31$ (after Katehi⁶ © IEEE 1990)

The typical polarization of most slotted-waveguide antennas is linear. However, several slotted-waveguide elements have been designed that produce circular polarization.^{7,8}

8.3 TAPERED AND FLARED SLOT ANTENNAS

Tapered slot antennas (TSAs) were first introduced in the late 1950s. It was then that Eberle et al⁹ produced a waveguide-fed, flared slot antenna for use in aircraft skins where conventional antennas could not be easily integrated. Gibson¹⁰ then developed the strip-line-fed, exponentially tapered slot antenna, which he called the *Vivaldi aerial*. His design was the first recognized TSA that showed symmetric *E*- and *H*-plane beamwidths, low sidelobes, and moderate gain.

Since then, with the increased interest in MIC (microwave integrated circuit) antennas for applications ranging from satellite communications to remote sensing, tapered slot antennas have been studied extensively, both empirically and theoretically. In general, the performance you can expect to achieve from a typical TSA includes

- Broadband operation
- Moderate gain
- Low sidelobes

These radiation characteristics make TSAs suitable elements for reflector feeds or as stand-alone antennas.

The current form of a typical TSA is planar in nature, etched on a thin, low-dielectric-constant microwave substrate, and fed by a stripline, finline, or slotline configuration, as shown in Figure 8-4. As a result of the typical substrates used in TSA designs, it is important to consider the possibility of the feedline itself radiating or coupling with the antenna element. Work at Harris Corporation has addressed this problem and has resulted in a bilateral stripline feed as one solution (see Figure 8-5).

TSAs are a type of traveling-wave antenna. Their operation is based on a traveling wave propagating along the surface of the antenna taper with a phase velocity less than the speed of light.¹³ Under this condition, endfire radiation results. Zucker reviews properties of traveling-wave antennas in Chapter 10.

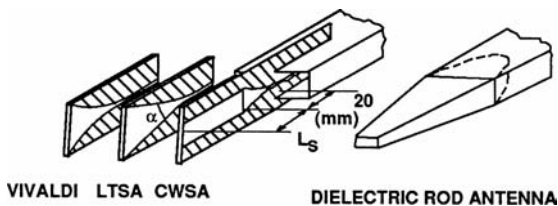


FIGURE 8-4 Three different types of endfire tapered slot antennas along with a dielectric-rod antenna shown for comparison (after Yngvesson et al^{11,12} © IEEE 1985 and © IEEE 1989)

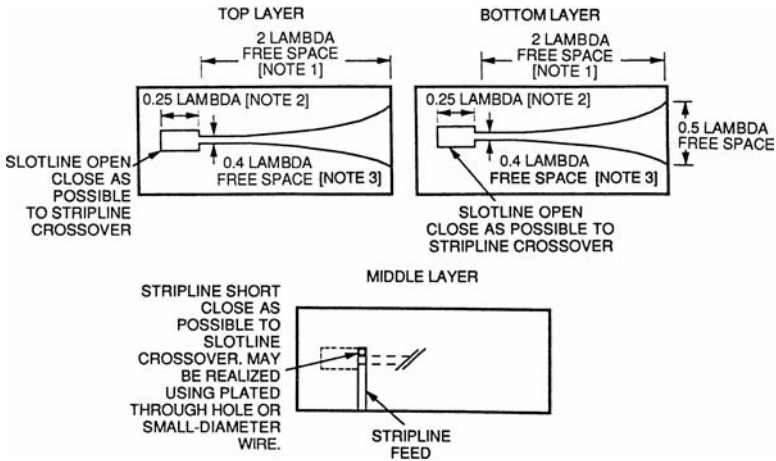


FIGURE 8-5 Stripline designs of the bilateral slotline-fed antenna. *Note 1:* Length of taper is proportional to gain; maximum gain approximately 10 dB. *Note 2:* Quarter wave calculated using dielectric constant of material. *Note 3:* May be narrower than 0.4 to obtain desired slotline impedance. Taper rate: $y = a^{kx}$, where $a = 0.5$ slotline width, $k = \text{constant}$, $x = \text{length of flare}$.

8.4 CAVITY-BACKED RECTANGULAR SLOT ANTENNAS

The electric field on the coaxially fed, cavity-backed rectangular slot (see Figure 8-6a) is neither sinusoidal nor complementary to a ribbon dipole antenna. This antenna is a cavity resonator, energized by the coaxial transducer, which radiates from the slot aperture. The field distribution in the slot, therefore, is dependent on the excitation of higher cavity modes as well as the principal mode (TE_{10}). The equivalent circuit of a cavity-backed slot antenna is shown in Figure 8-7b; the shunt conductance is the radiation conductance of the slot. The conductance of the cavity-backed resonant half-wave slot is half the open slot, free to radiate on both sides. That is, the shunt resistance is at least 800 rather than 400 Ω .¹⁴⁻¹⁶ The parallel susceptance shown in the equivalent circuit is the sum of the shunt susceptance of the slot radiator and the TE-mode susceptance of the cavity. The series-resonant circuit is the result of the energy stored in the TM modes in the cavity and feed structure.

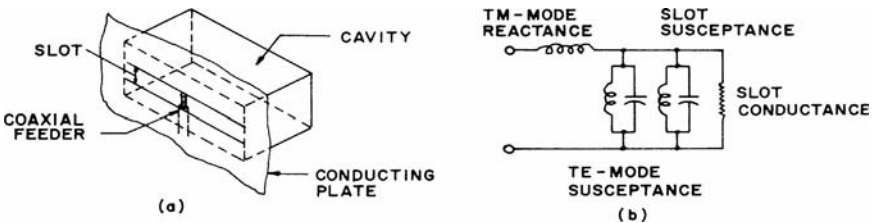


FIGURE 8-6 Cavity-backed rectangular slot: (a) pictorial representation; (b) equivalent circuit

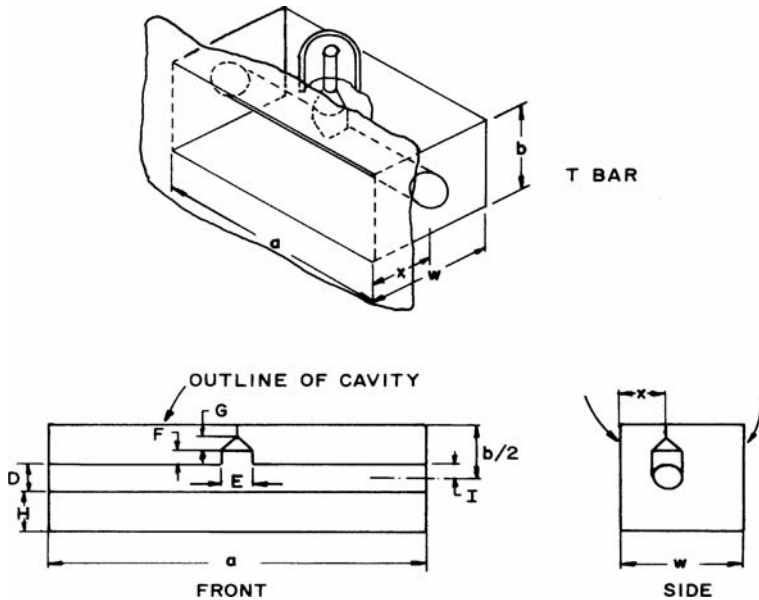


FIGURE 8-7 Cavity-backed T-bar-fed slot antenna. Typical dimensions shown in inches (millimeters) for a frequency range of 0.5 to 1.2 GHz: a , 12.00 (304.8); b , 4.00 (101.6); x , 3.25 (82.55); w , 6.75 (171.45); D , 2.25 (57.15); E , 0.75 (19.05); F , 0.63 (16.00); G , 0.19 (5.826); H , 0.25 (6.35).

To obtain the maximum radiation conductance, a sinusoidal distribution of electric field (magnetic current) must be generated. This distribution will be achieved when the energy stored in the cavity in the vicinity of the slot is primarily in the TE_{10} mode, i.e., by making the cavity dimensions big enough so that the dominant mode is above cutoff. For small cavities, edge loading, as in a highly capacitive slot, will improve the field distribution.

An important design parameter is the antenna Q , which is minimum when the stored energy is only in the dominant mode. The Q limits the inverse voltage-standing-wave-ratio (VSWR) bandwidth product; for a small cavity, it is

$$Q > \frac{3}{4\pi^2} \left(\frac{1}{V} \right)$$

where V is the volume of the cavity expressed in cubic free-space wavelengths. This minimum Q is realized when the series reactance is eliminated through efficient feed and cavity design. For the simple capacitive slot-loaded cavity shown in Figure 8-6a, higher TE and TM modes will be generated with attendant high Q .

A broadband cavity-backed antenna can be realized by using a T-bar feed,¹⁷ as shown in Figure 8-7. A flat T-bar instead of the illustrated circular cross section will generate the same impedance if its width is equal to the diameter D .

The nominal input impedance to the T-bar is 125 Ω (approximate). To achieve the available bandwidth, a broadband impedance transformer is needed between the 50- Ω coaxial transmission line and the T-bar junction.

The dimensions for a broadband flat T-bar covering the frequency range 0.5 to 1.2 GHz are shown in the diagram of Figure 8-7. The VSWR does not exceed 3:1 in the frequency band and is less than 2:1 over 90 percent of the band.

The resonant frequency can be modified by use of dielectric or ferrite loading within the cavity.¹⁸ The reduction in cavity volume and aperture size results in increased Q , smaller bandwidth, and lower efficiency.

A shallow ridged-cavity crossed-slot antenna¹⁹ has been developed for wide-angle coverage in the ultrahigh-frequency range. The VSWR is dependent on the slot width, slot length, and cavity depth at the low end of the band. The ridge parameters tune the antenna in the midband and high-band frequencies. The VSWR is less than 2.7:1 from 240 to 279 MHz and under 2.1:1 from 290 to 400 MHz for cavity dimensions of 33 by 33 by 4 in.

An experimental square cavity-slot antenna (half scale) with crossed slots cut along the diagonal dimensions of the cavity is shown in Figure 8-8. The cavity configuration, ridges, and crossed-slot arrangement are illustrated. The slots are excited by four symmetrically located feed probes near the center of the cavity. Each opposite pair is connected to a wideband 180° hybrid. For circular polarization, the input ports of the two 180° hybrids are connected to a wideband 90° hybrid. The 3-dB beamwidth varies from 120° at the low end of the band to about 40° at the high end.

An earlier narrowband crossed-slot antenna was developed with a cavity depth of 2 in.²⁰ A stripline version of the cavity-backed slot antenna has been designed.²¹ This design was used for many years as a low-profile antenna for rocket payloads.

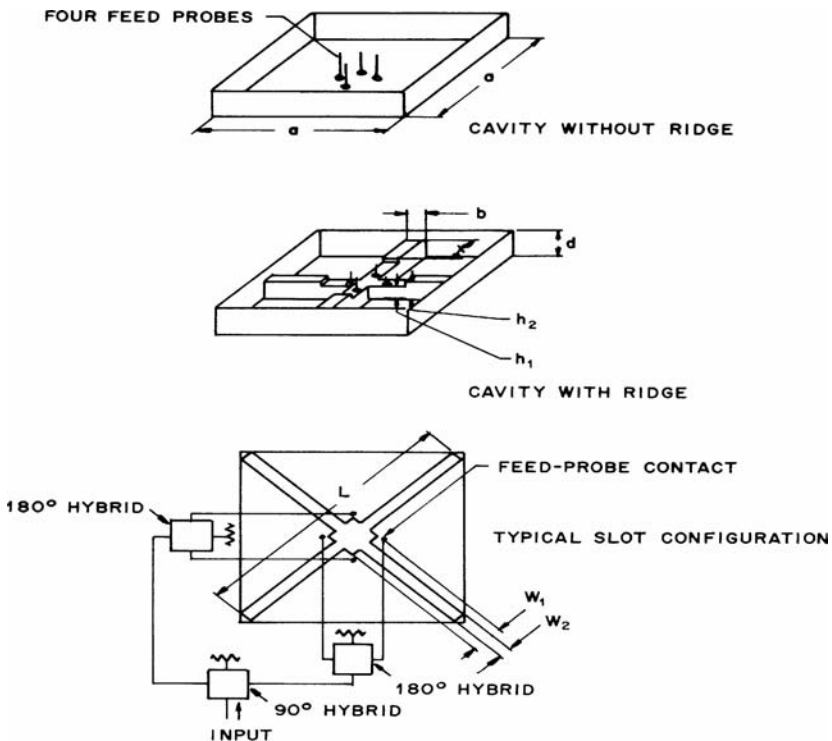


FIGURE 8-8 Cavity-slot configuration with circular polarization of a ridged-cavity crossed-slot antenna. Typical dimensions shown in inches (millimeters) for a frequency range of 480 to 800 MHz: a , 16.00 (406.4); d , 2.00 (50.8); b , 1.25 (31.75); W_1 , 2.50 (73.5); W_2 , 0.65 (16.51); h_1 , 1.50 (38.1); h_2 , 1.75 (44.45); L , 22 (558.8).

8.5 WAVEGUIDE-FED SLOT ANTENNAS

The waveguide opening onto a ground plane is commonly used in phased arrays and as a single element on curved as well as flat ground. An early analysis and experiment²² has been used extensively for checking various analytical methods. This early work has been extended to include a dielectric medium outside the slot,²³ as shown in Figure 8-9. Also given in Figure 8-9 are calculations and measurements that agree with those in Cohen et al.²² It is noted in Croswell et al.²³ that substantial energy is coupled into the dielectric slab in the form of surface waves. Another extension of this work includes dielectric plugs flush-mounted to the ground plane just inside the waveguide.²⁴ Dielectric plugs are very useful, but you must be careful of certain dielectric constants and plug thicknesses that can generate very significant higher-order aperture modes.

Similar analysis has been performed for circular waveguides.²⁵ The basic admittance of a typical TE₁₁-mode excited slot is given in Figure 8-10. This TE₁₁ excitation mode tends to couple less energy into the dielectric slab in the form of a surface wave. Hence the circular waveguide antenna was used in many early rocket reentry payload designs employing dielectric ablation materials.

The radiation characteristics of any annular slot (cut in an infinite ground screen) are identical with those of a complementary wire loop with electric and magnetic fields interchanged. In the case of the small slot, the radiation diagram is close to that of a small electric stub in the ground screen.

Consider a thin annular slot as shown in Figure 8-11. The polar axis of a spherical coordinate system being normal to the plane of the slot, the magnetic component of the radiated field is

$$H_{\phi} = \frac{aVe^{-jkr}}{120\pi\lambda r} \int_0^{2\pi} \cos(\phi - \phi') e^{jka \sin\theta \cos(\phi - \phi')} d\phi'$$

where a = radius of slot

V = voltage across slot

$K = 2\pi/\lambda$

For small values of a , that is, $a < \lambda/2\pi$,

$$H_{\phi} = j \frac{Ve^{-jkr}}{60r} \frac{A}{\lambda^2} \sin\theta \quad \text{A/m}$$

where $A = \pi a^2$ is the included area of the annular slot. The previous equation is valid for small slots of arbitrary shape.

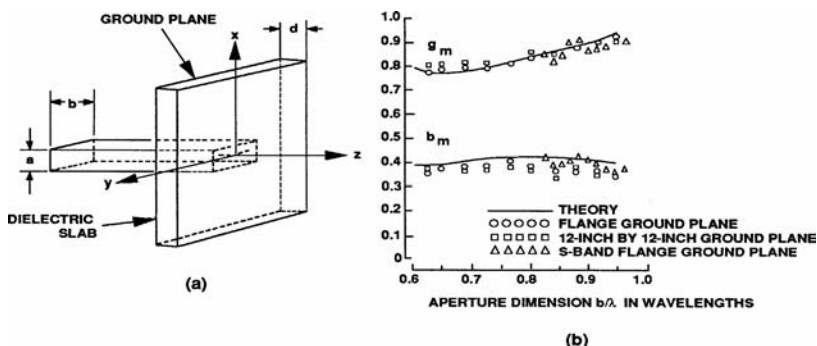


FIGURE 8-9 (a) Rectangular waveguide in a ground plane covered by a dielectric slab; (b) Effect of ground-plane size on the free-space admittance (after Croswell et al.²³ © IEEE 1967)

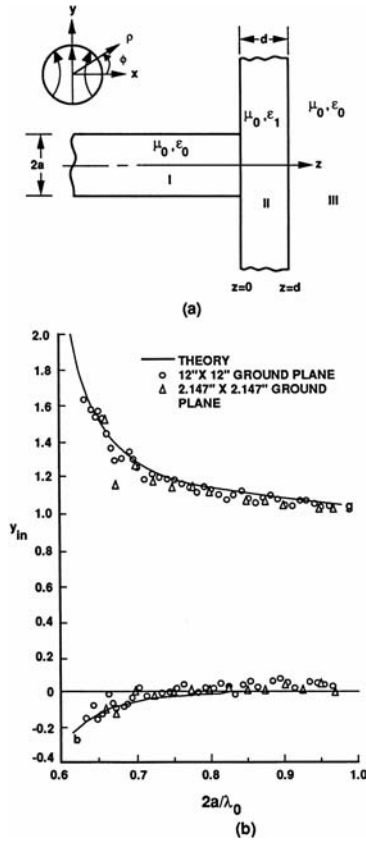


FIGURE 8-10 (a) Circular waveguide opening onto a dielectric-coated ground plane; (b) Input admittance of a circular waveguide radiating into free space (after Bailey and Swift²⁵ © IEEE 1968)

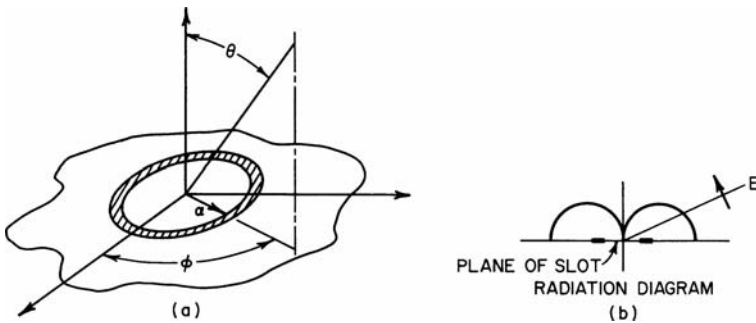


FIGURE 8-11 Annular slot: (a) coordinate system and (b) vertical-plane pattern for a small-diameter slot

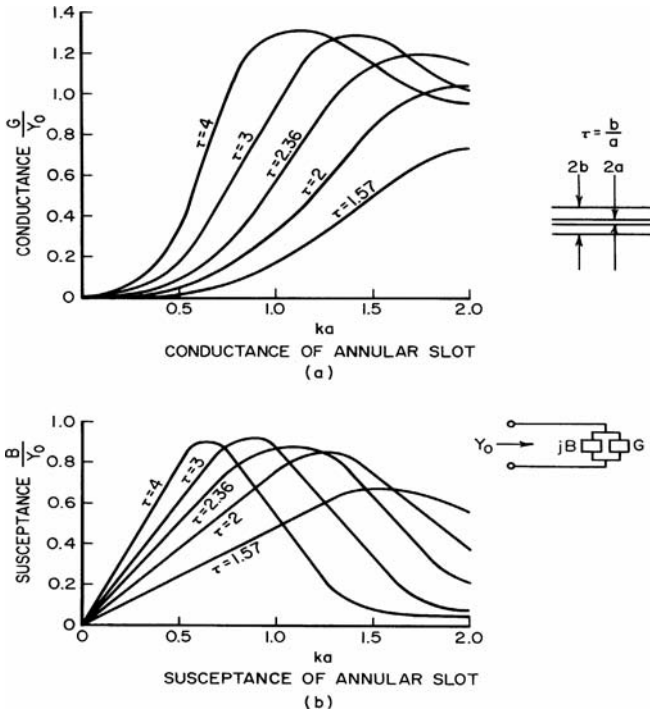


FIGURE 8-12 Admittance of a coaxial-fed annular slot in infinite ground plane

The integral in H_ϕ can be evaluated exactly as

$$H_\phi = j \frac{a V e^{-jkr}}{60r} J_1(ka \sin \theta)$$

where J_1 is the Bessel function of the first kind and the first order.

The radiation characteristics on a large but finite ground screen are closely approximated by the previous equations. There are slight perturbations because of edge effects that result in energy radiated into the shadow region plus modulation of the main radiation pattern.

The optimum excitation of an annular slot, i.e., least stored energy and lowest Q , results when the magnetic current distribution is uniform around the slot. One method for obtaining this result is to feed the annular slot by a coaxial transmission-line structure that has the same inner and outer diameters as the annular structure. Figure 8-12 consists of graphs of the conductance and susceptance in the plane of the aperture relative to the characteristic admittance of the feedline as a function of the radian length ka of the inner radius a . It is seen that the slot is at all times nonresonant and has a capacitive susceptance.

8.6 SLOT ANTENNAS ON FINITE AND CURVED GROUND PLANES

The annular slot is commonly used on aircraft as a UHF antenna because it produces a pattern similar to a short vertical dipole. Early research on annular slot antennas for the space

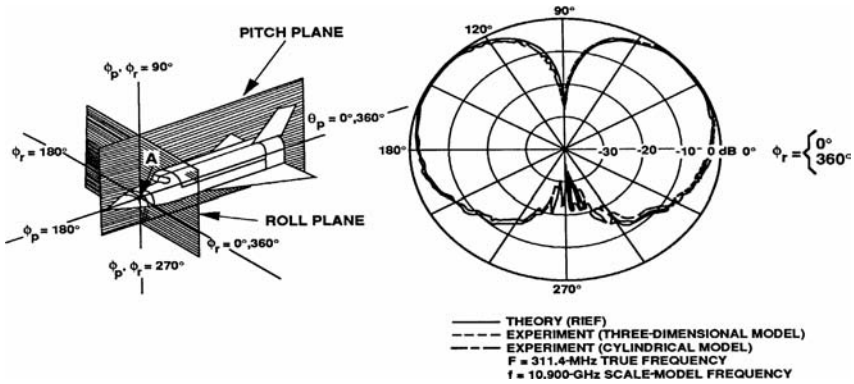


FIGURE 8-13 Roll-plane pattern of annular-slot antenna mounted on a model of the space shuttle (after Jones and Richmond²⁶ © IEEE 1974)

shuttle has been conducted and reported.²⁶ This analysis is an extension of the moment method to semi-infinite strips in an integral equation formulation. It was determined that for many cross sections through the body shape, the ϕ plane pattern was similar for either a two- or three-dimensional complex shape, as shown in Figure 8-13.

Additional extensive research has been performed for slot antennas on finite and curved ground planes. The early work done by Jones and Richmond²⁶ using integral equation solutions to determine the E -field distribution has been extended to apply to patches of surface instead of strips. This type of EM analysis, commonly referred to as the *method of moments (MOM)*, however, is limited to small surface volumes due to limitations in computers. An asymptotic method, the *geometrical theory of diffraction (GTD)*, has been developed that is not limited by the ground-plane size in wavelengths.

The GTD is an extension of *geometrical optics (GO)*, which is demonstrated in Figure 8-14. If you consider GO *only*, there are no fields in region III past the shadow boundary. Actually, the diffracted fields do exist and can be computed by GTD.²⁷

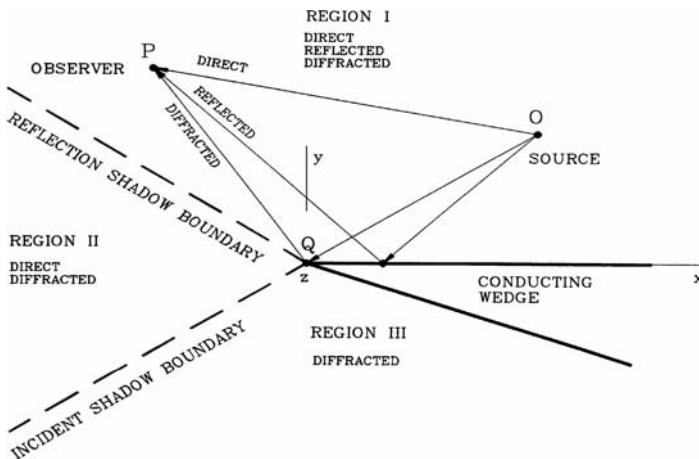


FIGURE 8-14 Two-dimensional electric conducting wedge and field regions (after Balanis²⁷ © Harper & Row 1982)

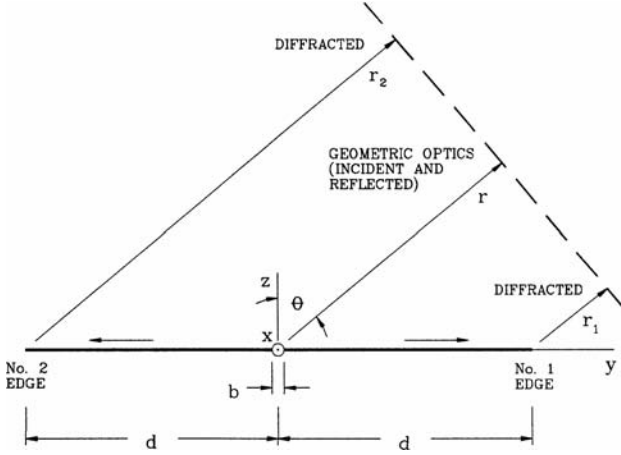


FIGURE 8-15 Aperture geometry in the principal-plane E -plane ($\phi = \pi/2$) (after Balanis²⁷ © Harper & Row 1982)

A simple example of the use of GTD is depicted in Figure 8-15. Following Balanis,²⁷ the total fields in all regions can be summarized as

$$Et = E_{go}(r, \theta) + E_1^d(\theta) + E_2^d(\theta)$$

where

$$E_{go}(r, \theta) = E_0 \left[\frac{\sin\left(\frac{ka}{2} \sin \theta\right)}{\frac{ka}{2} \sin \theta} \right] \frac{e^{-jkr}}{r} \quad 0 \leq \theta \leq \frac{\pi}{2}$$

$$E_1^d(\theta) = \frac{E_0 \sin\left(\frac{ka}{2}\right)}{\frac{ka}{2}} V_B^{i,r}(d, \beta = \theta + \pi/2, 2) e^{jkd \sin \theta} \frac{e^{-jkr}}{r}$$

$$E_2^d(\theta) = \frac{E_0 \sin\left(\frac{ka}{2}\right)}{\frac{ka}{2}} V_B^{i,r}(d, \beta, 2) e^{jkd \sin \theta} \frac{e^{-jkr}}{r}$$

where

$$\beta = \begin{cases} \pi/2 - \theta & 0 \leq \theta \leq \pi/2 \\ 5\pi/2 - \theta & \pi/2 \leq \theta \leq \pi \end{cases}$$

and $V_B^{i,r}$ is the wedge diffraction coefficient. A typical result for a slot on a finite ground plane is given in Figure 8-16.

This work has been extended to slots opening to both square and elliptical cylinders,²⁸ as shown in Figure 8-17. Note that the patterns are qualitatively similar except for the ripple caused by diffractions at the edges of the rectangular cylinder. Further applications of GTD to three-dimensional surfaces for slot-antenna pattern calculations have been extensively pursued.

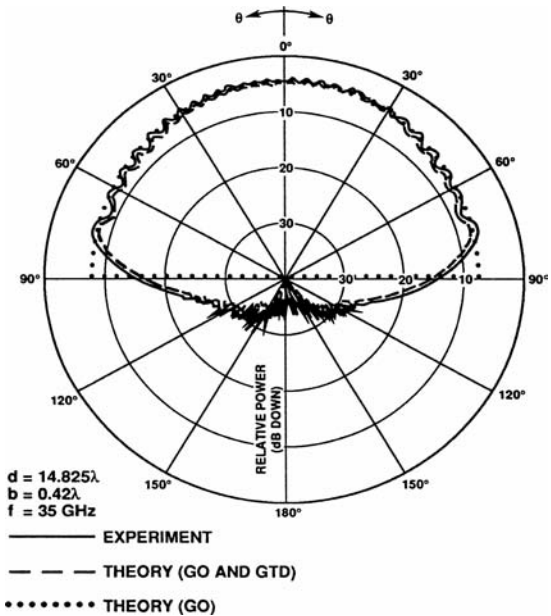


FIGURE 8-16 Principal *E*-plane amplitude patterns of an aperture antenna mounted on a finite-size ground plane (after Balanis²⁷ © Harper & Row 1982)

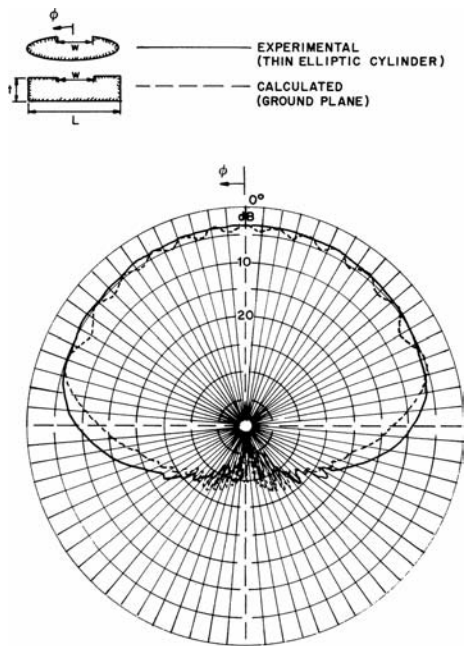


FIGURE 8-17 Radiation patterns of a thin elliptical cylinder and finite ground plane (TEM mode)

8.7 SLOT ANTENNAS ON CYLINDERS, CONES, AND SPHERES

A common design problem for spacecraft, aircraft, and rocket antennas is to determine the pattern and impedance of such antennas when located on curved dielectric-coated surfaces where the conformal dielectric is a radome, a reentry ablator, and/or an absorber. To obtain first-order results, it has been found useful to consider the radiation characteristics of slot antennas on cylinders, cones, and spheres.

The best early summary for slots on cylinders is a book by Wait.²⁹ Accurate calculations of the patterns of a slot on an uncoated cylinder have been published by Knop and Battista,³⁰ a few of which are given in Figure 8-18; note that $c = ka$. If a coating is applied to the cylinder surface, a series of azimuthal surface waves can be excited that can produce major ripples in the pattern depending on dielectric constant and thickness. An estimate of such modes can be made using equations in Croswell et al.²³ Measured patterns of a coated azimuthal slot on a cylinder are available.³¹ Similar effects were noted for axial slots on dielectric-coated cylinders (unpublished notes).

The admittance of a slot on a coated cylinder also has been computed.³² The admittance is about the same as that on a flat coated ground plane with

$$ka = \frac{2\pi a}{\lambda} \geq 3 \text{ to } 4$$

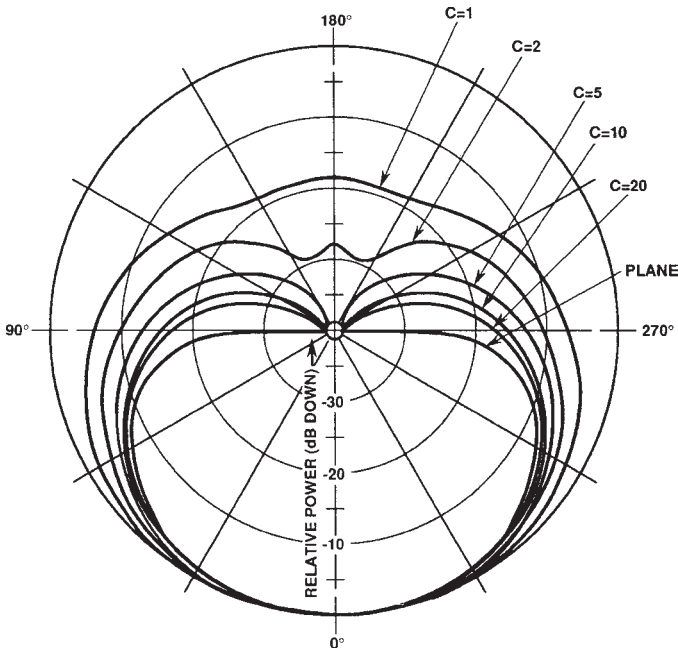


FIGURE 8-18 Equatorial-plane power patterns for a half-wavelength slot on an uncoated cylinder (after Knop and Battista³⁰ © IEEE 1961)

Interestingly, about $2ka$ terms are required to compute a convergent value of conductance. However, the susceptance requires about 75 to 100 terms regardless of ka . The susceptance in this formulation represents the near field of the slot.

A major problem on early satellites was to design an antenna having a quasi-isotropic radiation pattern on spherical-shaped surfaces with large values of ka . Early data were available for small spheres, but patterns for larger spheres excited by a parallel-plate waveguide were first considered by Bugnolo,³³ who took great care to make precise measurements. More practical forms of the large spherical antenna excited by a slotted waveguide and a modified parallel-plate waveguide have been repeated.^{31,34}

Another slot antenna of great interest for missiles and aircraft is the slotted cone. This problem is dominated by the tip diffraction.³⁵ Additional results for a slot on a cylinder fed by a cavity are given in Swift et al.³⁶ Antennas similar to this were mounted on sphere-tipped cones and used for telemetry antennas on reentry payloads.

REFERENCES

1. H. G. Booker, "Slot Aerials and Their Relation to Complementary Wire Aerials," *J. IEE (London)*, part IIIA, vol. 93 (1946): 620–626.
2. Y. L. Purnam, B. Russell, and W. Walkinshaw, "Field Distributions Near a Center Fed Half-Wave Radiating Slot," *J. IEE (London)*, part III, vol. 95 (July 1948): 282–289.
3. A. A. Oliner, "The Impedance Properties of Narrow Radiating Slots in the Broad Face of Rectangular Waveguide—Part I: Theory; Part II: Comparison with Experiment," *IRE Trans. Antennas Propagat.*, vol. AP-15 (1967): 594–598.
4. A. F. Stevenson, "Theory of Slots in a Rectangular Waveguide," *J. Appl. Phys.*, vol. 19 (1948): 24–38.
5. M. C. Bailey, "The Impedance Properties of Dielectric-Covered Narrow Radiating Slots in the Broad Face of a Rectangular Waveguide," *IEEE Trans. Antennas Propagat.*, vol. AP-18, no. 5 (September 1970): 596–603.
6. P. B. Katehi, "Dielectric-Covered Waveguide Longitudinal Slots with Finite Wall Thickness," *IEEE Trans. Antennas Propagat.*, vol. AP-38, no. 7 (July 1990): 1039–1045.
7. A. J. Simmons, "Circularly Polarized Slot Radiators," *IRE Trans. Antennas Propagat.*, vol. AP-5, no. 1 (January 1957): 31–36.
8. J. A. Ajioka, D. M. Joe, and J. L. McFarland, "Slot Radiators in Septated Waveguide," *IEEE Trans. Antennas Propagat.*, vol. AP-32, no. 3 (March 1984): 247–251.
9. J. W. Eberle, C. A. Lewis, and D. McCoy, "The Flared Slot: A Moderately Directive Flush Mounted Broadband Antenna," *IRE Trans. Antennas Propagat.*, vol. AP-8 (1960): 461–468.
10. P. J. Gibson, "The Vivaldi Aerial," *Proc. 9th European Microwave Conf.* (1979): 101–105.
11. K. S. Yngvesson et al, "Endfire Tapered Slot Antennas on Dielectric Substrates," *IEEE Trans. Antennas Propagat.*, vol. AP-33, no. 12 (December 1985): 1392–1400.
12. K. S. Yngvesson et al, "The Tapered Slot Antenna—A New Integrated Element for Millimeter Wave Applications," *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 2 (February 1989): 365–374.
13. R. Janaswamy and D. H. Schaubert, "Analysis of the Tapered Slot Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-35, no. 9 (September 1987): 1058–1065.
14. C. R. Cockrell, "The Input Admittance of the Rectangular Cavity-Backed Slot Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 3 (May 1976): 288–294.
15. S. A. Long, "Experimental Study of the Impedance of Cavity-Backed Slot Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-23, no. 1 (January 1975): 1–7.
16. S. A. Long, "A Mathematical Model for the Impedance of the Cavity-Backed Slot Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-25, no. 6 (November 1977): 829–833.

17. E. H. Newman and G. A. Thiele, "Some Important Parameters in the Design of T-Bar Fed Slot Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-23, no. 1 (January 1975): 97–100.
18. A. T. Adams, "Flush Mounted Rectangular Cavity Slot Antennas—Theory and Design," *IEEE Trans. Antennas Propagat.*, vol. AP-15, no. 3 (May 1967): 342–351.
19. H. E. King and J. L. Wong, "A Shallow Ridged-Cavity Crossed-Slot Antenna for the 240–400 MHz Frequency Range," *IEEE Trans. Antennas Propagat.*, vol. AP-23, no. 5 (September 1975): 687–689.
20. C. A. Lindberg, "A Shallow-Cavity UHF Crossed-Slot Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-17, no. 5 (September 1969): 558–563.
21. T. A. Milligan, *Modern Antenna Design* (New York: McGraw-Hill, 1985): 96–98.
22. M. H. Cohen, T. H. Crowley, and C. A. Levis, "The Aperture Admittance of a Rectangular Waveguide Radiating into Half-Space," Antenna Lab, Ohio State University Research Foundation, Columbus, Ohio, Rept. 339–22, Cont. USAF W33-038-ac-21114 (November 1951).
23. W. F. Croswell, R. C. Rudduck, and D. M. Hatcher, "The Admittance of a Rectangular Waveguide Radiating into a Dielectric Slab," *IEEE Trans. Antennas Propagat.*, vol. AP-15, no. 5 (September 1967): 627–633.
24. C. T. Swift and D. M. Hatcher, "The Input Admittance of a Rectangular Aperture Loaded with a Dielectric Plug," NASA, TND-4430 (April 1968).
25. M. C. Bailey and C. T. Swift, "Input Admittance of a Circular Waveguide Aperture Covered by a Dielectric Slab," *IEEE Trans. Antennas Propagat.*, vol. AP-16, no. 4 (July 1968): 386–391.
26. J. E. Jones and J. H. Richmond, "Application of an Integral Equation Prediction of Space Shuttle Annular Slot Antenna Radiation Patterns," *IEEE Trans. Antennas Propagat.*, vol. AP-22, no. 1 (January 1974): 109–111.
27. C. A. Balanis, *Antenna Theory, Analysis, and Design* (New York: Harper & Row, 1982).
28. C. A. Balanis and L. Peters, Jr., "Radiation from TE_{10} Mode Slots on Circular and Elliptical Cylinders," *IEEE Trans. Antennas Propagat.*, vol. AP-17, no. 4 (July 1969): 507–513.
29. J. R. Wait, *Electromagnetic Radiation from Cylindrical Structures* (New York: Pergamon Press, 1959).
30. C. M. Knop and A. R. Battista, "Calculated Equatorial Plane Radiation Patterns Produced by a Circumferential Slot on a Cylinder," *IEEE Trans. Antennas Propagat.*, vol. AP-9, no. 5 (September 1961): 498–499.
31. W. F. Croswell, C. M. Knop, and D. M. Hatcher, "A Dielectric-Coated Circumferential Slot Array for Omnidirectional Coverage at Microwave Frequencies," *IEEE Trans. Antennas Propagat.*, vol. AP-15, no. 6 (November 1967): 722–727.
32. W. F. Croswell, G. C. Westrick, and C. M. Knop, "Computations of the Aperture Admittance of an Axial Slot on a Dielectric Coated Cylinder," *IEEE Trans. Antennas Propagat.*, vol. AP-20, no. 1 (January 1972): 89–72.
33. D. S. Bugnolo, "A Quasi-Isotropic Antenna in the Microwave Spectrum," *IRE Trans. Antennas Propagat.*, vol. 10, no. 4 (1962): 377–383.
34. W. F. Croswell and C. R. Cockrell, "An Omnidirectional Microwave Antenna for the Spacecraft," *IEEE Trans. Antennas Propagat.*, vol. AP-17, no. 4 (July 1969): 459–466.
35. D. C. Pridmore-Brown and G. E. Stewart, "Radiation from Slot Antennas on Cones," *IEEE Trans. Antennas Propagat.*, vol. AP-20, no. 1 (January 1972): 36–39.
36. C. T. Swift, T. G. Campbell, and H. Hodara, "Radiation Characteristics of a Cavity-Backed Cylindrical Gap Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-17, no. 4 (July 1969): 467–476.

Chapter 9

Waveguide Slot Antenna Arrays

Roland A. Gilbert

BAE Systems, Inc.

CONTENTS

9.1 INTRODUCTION	9-2
9.2 WAVEGUIDE SLOT RADIATORS	9-2
9.3 SLOTTED WAVEGUIDE ARRAY DESIGNS	9-5
9.4 COMPUTATION METHODS FOR SLOT CHARACTERIZATION	9-18
9.5 DESIGN PARAMETERS	9-22
9.6 POWER-HANDLING CAPABILITIES OF SLOTTED-WAVEGUIDE ARRAYS	9-32
9.7 TOLERANCE AND FABRICATION TECHNIQUES	9-34

9.1 INTRODUCTION

Geometric simplicity, efficiency, polarization purity, conformal installation, and ability to radiate broadside beams and vertically polarized E -plane beams at very near grazing angle above a ground plane make slot-antenna arrays ideal solutions for many radar, communications, and navigation applications. Especially today with the desire to make antennas for aircraft as low profile as possible to reduce drag and conserve fuel, slot antenna arrays can be positioned above wings and on top of the fuselage while having the capability to look toward the horizon. Classical slot arrays are depicted as narrow nonconductive slits etched or milled into the host metallic ground plane. They are characterized by the methods used to feed or excite the slots. Without what is behind the slot opening being seen, slot radiators appear similar on the surface of the ground plane. Narrow conformal slots tend to be narrowband (< 5 percent f_0) and have high cross-polarization isolation when operating near their resonant frequency. Wider slots can exceed an octave bandwidth given a well-matched feed. However, polarization purity is usually not as good as with the narrower slots. Conformal slot arrays are generally limited in bandwidth because the array lattice spacing has to be large enough to accommodate the waveguide and feed structures behind the slots without creating grating lobes.

Conformal slot elements can be fed in a variety of ways: (1) tapping into a transmission line such as a waveguide, (2) coupling to a resonant cavity, and (3) feeding them directly with voltage sources across the slots. Each method has some impact on the radiation performance and operating bandwidth of the slot radiators and hence the array. This chapter focuses on the first approach, which is the transmission-line method whereby slots are cut along a waveguide to couple energy in a controlled manner to slots that radiate. Waveguide-fed array systems are either traveling-wave or standing-wave approaches. Because waveguides are dispersive transmission lines, the array excitation along the waveguide has a differential phase relationship between elements that changes with frequency, causing the array beam to scan. For fixed beam (nonscanning) arrays, the waveguide is converted into a resonant, standing-wave structure. Before the advent of broadband MMIC T/R module technology, waveguide-fed slot arrays were most common in microwave radar applications. Today corporate feed networks offer a greater flexibility to excite slot arrays over a broader bandwidth. They are most effective when T/R modules, comprised of amplifiers, attenuators, and phase shifters or time delay devices, are connected in series with each array element. The slot elements can be phased independently to scan a beam anywhere in a hemisphere above their host ground plane. However, such large phased arrays that have independent control of every radiator are still prohibitively expensive. Therefore array architectures are employed that utilize hybrid scan approaches where electronic scanning is used in one plane, and either mechanical or frequency scanning is used in the other plane. Thus waveguide slot arrays can still provide a very cost-effective solution to fulfilling many high performance array needs.

9.2 WAVEGUIDE SLOT RADIATORS

The radiating elements of a waveguide slot array are an integral part of the feed system, which is the waveguide itself. This simplifies the design since baluns or matching networks are not required. A familiarization with the modal fields within a waveguide is necessary to understand where to place slots so that they are properly excited. Narrow slots that are parallel to waveguide wall currents do not radiate. However, when a slot is cut into a waveguide wall and it interrupts the flow of current, forcing it to go around the slot, power is coupled from the waveguide modal field through the opening to free space. To have good control of the excitation of a linear slot array, it is recommended that the waveguide

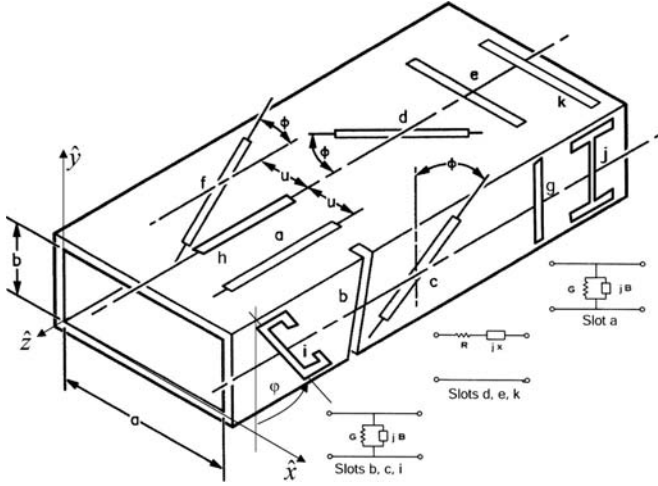


FIGURE 9-1 Slots cut in the walls of a rectangular waveguide. Slot *g* does not radiate because the slot is lined up with the direction of the sidewall current. Slot *h* does not radiate because the transverse current is zero there. Slots *a*, *b*, *c*, *i*, and *j* are shunt slots because they interrupt the transverse currents (J_x , J_y) and can be represented by two-terminal shunt admittances. Slots *e*, *k*, and *d* interrupt J_z and are represented by series impedance. Slot *d* interrupts J_x , but the excitation polarity is opposite on either side of the waveguide centerline, thus preventing radiation from that current component. Both J_x and J_z excite slot *f*. A Π - or T -impedance network can represent it.

only operate in a single mode, preferably the lowest mode. When a waveguide as shown in Figure 9-1 is excited with a TE_{10} mode and the ends are terminated in a matched impedance, the fields are given by Eq. 9-1.

$$\begin{aligned}
 H_x &= \frac{-\beta_z}{\omega\mu} E_o \sin(\beta_x x) e^{-j\beta_z z} \\
 E_y &= E_o \sin(\beta_x x) e^{-j\beta_z z} \\
 H_z &= \frac{j\beta_x}{\omega\mu} E_o \cos(\beta_x x) e^{-j\beta_z z}
 \end{aligned} \tag{9-1}$$

where

$$\begin{aligned}
 \beta_x &= \pi/a \\
 \beta_z &= \sqrt{k^2 - \beta_x^2} = \frac{2\pi}{\lambda_g} \\
 k &= \frac{2\pi}{\lambda} = \frac{\omega}{c} \\
 \lambda_g &= \frac{\lambda}{\sqrt{\lambda_c^2 - \lambda^2}} \\
 \lambda_c &= 2a
 \end{aligned} \tag{9-2}$$

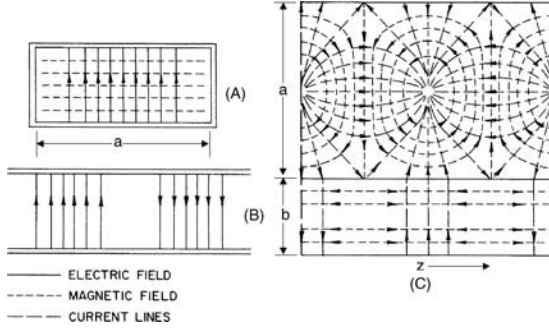


FIGURE 9-2 Surface-current distribution for rectangular waveguide propagating TE_{10} mode: (a) Cross-sectional view shows E - and H -fields. (b) Longitudinal view shows polarity of E -field along waveguide. (c) Surface views show top and sidewall currents and H -field.

The currents J along the inner waveguide wall surfaces are proportional to H .

$$\vec{J} = \hat{n} \times \vec{H} \quad (9-3)$$

The currents along the top and side walls are shown in Figure 9-2. At the top inner wall surface ($y = b$),

$$J_x = -j \frac{\beta_x}{\omega \mu} E_o \cos(\beta_x x) e^{-j\beta_z z} \quad (9-4)$$

$$J_z = -\frac{\beta_z}{\omega \mu} E_o \sin(\beta_x x) e^{-j\beta_z z}$$

On the bottom inner wall surface ($y = 0$),

$$\vec{J}_{\text{bottom}} = -\vec{J}_{\text{top}} \quad (9-5)$$

On the inner sidewalls ($x = 0$, $x = a$), there is only a \hat{y} -directed current with the same phase.

$$J_y = -j \frac{\beta_x}{\omega \mu} E_o e^{-j\beta_z z} \quad (9-6)$$

Rotating the slot with respect to a peak current direction can control the power coupled to a slot. For example, slot e couples maximum power, while the power is proportional to $\sin^2 \phi$ for slots d and c . Another way to control coupled power is to take advantage of the natural field intensities within the waveguide by locating the slots accordingly. For example, J_x is a null at the center of the surface wall and varies sinusoidally as you approach the edge. Therefore, by offsetting longitudinal slots such as slot a from the center of the waveguide, the power coupled to the slots can be adjusted. The ability to control the excitation of slots in a linear waveguide is important in order to design arrays with tapered sidelobes. Moreover, depending on how the array is fed, the coupling of the waveguide to the slots must vary progressively down the length of the waveguide if the first elements are not to radiate all the power, with little power left for the remaining elements.

Slots are classified by their shapes, locations in the rectangular waveguide, and how they are arrayed. Slots are usually about $\lambda/2$ long at the center frequency of their operating bandwidths. For slots located on the wider waveguide surfaces ($a = \lambda_c/2$), there is plenty of room for offset and rotated slots. For slots cut into the side walls, where the rotation angle ϕ cannot be large and $b < a/2$, there is usually not enough room for $\lambda/2$ slots. The slots are either extended or “wrapped around” into the adjoining surfaces (e.g., slot *b* in Figure 9-1) or the slots are end loaded (e.g., *C* and *I* slots) to resonate them. Wrapped-around slots are not as desirable in conformal planar array configurations because the elements have to be slightly elevated above the ground plane or gaps or spacers provided between waveguides. Complicated slot designs are also more costly to manufacture.

Slots are usually represented as rectangular shapes, which simplifies analysis. Unless they are etched on a metallized substrate, narrow slots on the broad wall of a waveguide are usually fabricated by a milling process in which rounded ends are a natural output of the process. Wider slots can be manufactured with straight ends but with rounded corners. The impact of rounded ends is small but does change the impedance of the slot and hence its resonant frequency.

9.3 SLOTTED WAVEGUIDE ARRAY DESIGNS

A slotted waveguide array is, by its very nature, a linear or *stick* array. Planar slotted waveguide arrays are comprised of multiple stick arrays placed side by side, as shown in Figure 9-3. The mutual coupling between slots in a planar array is high, especially between slots along the *E*-plane, such that dimensional and positional modifications to the slots are needed to compensate for resonant frequency changes. Most slot array designs are for fixed beam or mechanically scanned applications where the array is mounted on a gimbaled pedestal to steer the beam. Low loss beamforming networks to feed the arrays can also be constructed out of waveguides. Sum/difference waveguide components, such as magic-*T*s,

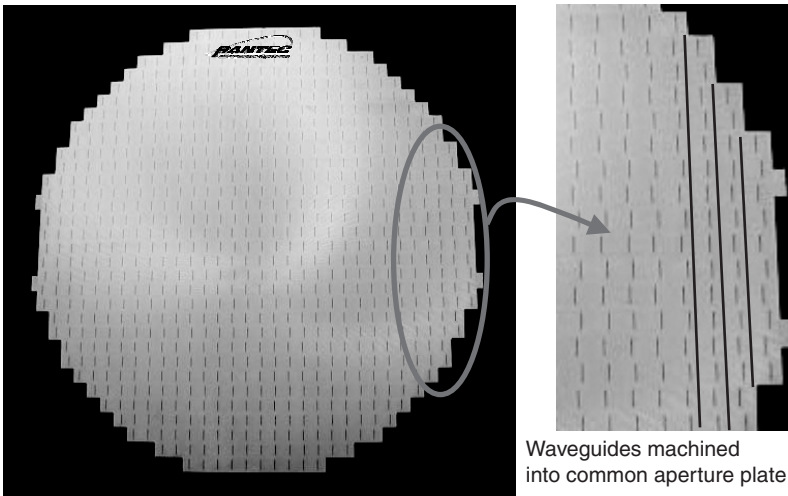


FIGURE 9-3 Planar longitudinal shunt-slot array. Adjacent waveguides are indicated at right.

simplify the implementation of monopulse networks in planar arrays for use in radar or satellite tracking applications. Electronic scanning is possible along the axis transverse to the waveguide stick subarrays. Phase shifters and attenuators are included in the network that feeds signals to each stick subarray. However, the active impedance of the array is highly affected with scan angle, especially in the case of longitudinal slots where the polarization is such that the E -plane is the scan plane. Hence the array design must be optimized with scanning requirements in mind.

If a TEM transmission line fed slot elements, the beam would radiate at a fixed angle as a function of the delay between elements, independent of changes in frequency. Waveguides are dispersive and the group velocity or delay between slots varies with frequency. Therefore the beam radiated by a slot array scans along the longitudinal axis of the waveguide as the frequency changes. The group delay is not a linear function of frequency, and it increases rapidly as the frequency decreases to near cutoff frequency. For a fixed beam array, this is a feature, or a problem as the case might be, that limits the bandwidth over which the array can be used. Frequency scanned radars make use of this feature to scan a pencil beam in a plane by “chirping” the signal. The array is either mechanically scanned in the other plane or electronically scanned through a network with phase shifters.

Waveguide slot arrays are classified into two groups: (1) standing-wave arrays and (2) traveling-wave arrays. The standing-wave arrays have elements spaced $\lambda_g/2$ and radiate a beam broadside to the waveguide. The fields repeat in a waveguide every $\lambda_g/2$ but are of opposite phase. Therefore the slots are placed in a $+/-$ configuration so that they are all fed in phase. Because of dispersion in a waveguide, the bandwidth around the center operating frequency cannot deviate by more than a few percent without causing rapid deterioration of the beam pattern and sidelobe levels, especially for the standing-wave array. Standing-wave arrays can be fed either at one end of the waveguide with the other end terminated in a matched load or short circuit, or at the center of the waveguide with matched load or short circuit terminations at the waveguide ends. Short circuit terminations provide for a more efficient array since the reflected wave from the waveguide ends can be phased with the incident wave. This allows for a higher power handling capability. If a slightly larger bandwidth is desired, matched terminations minimize reflected waves that potentially could cause the array to radiate another beam in the opposite direction.

Traveling-wave arrays are used in applications where the direction of the main beam is pointed at angles that are not broadside to the waveguide wall or where frequency scanning is desired. Inter-element spacing does not have to be the same between the elements, and $\lambda_g/2$ spacing is particularly avoided. In designing these arrays it is important to have wide-band terminations with a very low VSWR to prevent the formation of reflected or backward waves. Backward waves appear to originate from the opposite end of the waveguide and excite the slots to produce unwanted beams in the backward direction. Traveling-wave arrays can only be fed from the ends of the waveguide. To maximize the impedance bandwidth of the array, the slot elements are designed to be resonant at their center operating frequency, hence they are called *resonant* slots. This should not be confused with resonant arrays discussed earlier which, unlike traveling-wave arrays, have a standing-wave condition within the waveguide.

Initial Slot Array Design Considerations

One advantage of waveguide slot arrays is that not only the radiating aperture but also the feed network can be made from waveguides. Waveguides can couple energy to other waveguides in a very precise manner. They are also the preferred low loss transport medium to feed arrays at X-band frequencies and above. Several methods can be used to manufacture the radiating aperture. One popular way is to mill waveguide troughs into a solid piece of aluminum and

then to machine the slots. Therefore waveguide width can be made to any custom dimension that is not in cutoff at the operating frequency. This allows design flexibility in the selection of standard waveguide sections for the feed network in the rear of the array aperture. However, be careful to keep inter-element spacing less than a free-space wavelength λ_0 to prevent the appearance of grating lobes. For scanning arrays, the spacing between elements should be closer to $\lambda_0/2$ in the plane of scan. For resonant slotted-waveguide arrays, the element spacing along the waveguide must be equal to half the waveguide wavelength ($\lambda_g/2$) since all the element excitations must be in-phase. For nonresonant slotted-waveguide arrays, inter-element spacing is slightly larger or smaller than $\lambda_g/2$. The waveguide wavelength is larger than the free-space wavelength for the same frequency.

Designing slot arrays is similar to designing any other fixed beam or phased arrays. Once the array requirements are determined, such as gain, sidelobe level, beamwidth, bandwidth, polarization, input VSWR, scan impedance, cross-polarization level, power-handling capability, etc., then the array element excitations are calculated. Most array design codes available that consider mutual coupling between array elements utilize the infinite array approach, which is numerically efficient. However, it does not include array edge and truncation effects. These must be computed by low-frequency methods such as MOM or FEM. Because slots will have different excitations although they have the same resonant frequency, they will have different offset positions from the centerline of the waveguide to achieve their resonant conductances and/or resistances, as in the case of longitudinal shunt slots. Series and edge-wall slots are rotated along the waveguide axis. The element layout is not exactly periodic. Hence slot placement and orientation must be approximated. Once an array excitation is obtained that includes compensation for edge and radome effects, then the element coupling can be readjusted. Since the elements have been moved, a last performance prediction is made to verify that performance will be as desired. This design process is outlined in Figure 9-4.

To optimize bandwidth performance, the slots are designed to be resonant at midband for either resonant or nonresonant arrays. The power coupled from the waveguide and

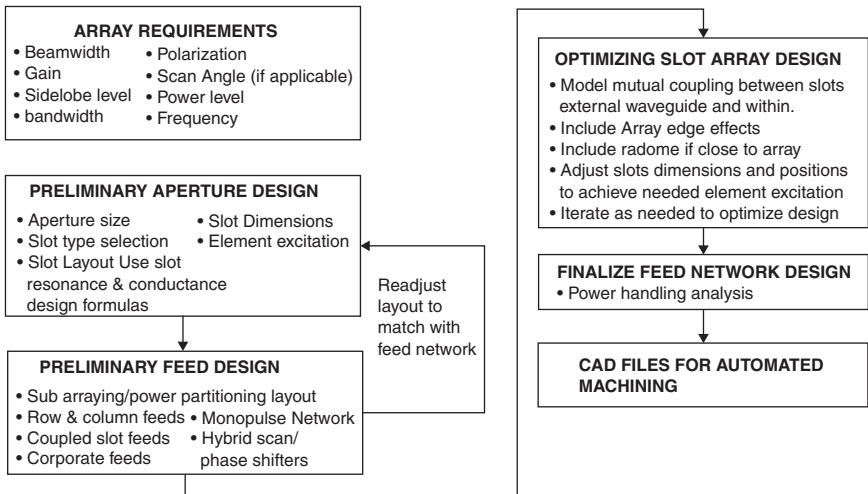


FIGURE 9-4 Slot array design process. Once the slot element excitations are determined to meet the array beamwidth, gain, and sidelobe requirements, the next step is to lay out the slots onto waveguide. Considerations must be given to the beamformer design. One or more design iterations will be needed to optimize the design.

radiated by the slot is proportional to slot conductance or resistance. If P_i is the fraction of the total waveguide power radiated by slot i , the normalized shunt-slot conductance or series-slot resistance is given by

$$g = 2P/V^2 \quad r = 2P/I^2 \quad (9-7)$$

where V or I are respectively the voltage or current across the slot. The transmission-line or scattering-matrix theories can be used to determine V and I and, therefore, g and r . Methods to achieve certain slot characteristics are described in later sections. First let us examine the differences between the various types of slot arrays.

Resonant Slotted-Waveguide Array Design

Linear and planar resonant slotted-waveguide arrays are used in many applications, especially radar. The advantage of using a planar slotted waveguide array over a parabolic reflector, for example, is that the array is thinner, the beam and sidelobes can be shaped, and two-axis monopulse is more easily implemented. The operating principle of resonant slotted-waveguide arrays is that the slot elements are excited by a standing wave set up within the waveguide, and hence all have the same phase. The impedance (admittance) repeats itself in a transmission line at every $\lambda_g/2$. If the impedances (admittances) of series (shunt) slots are real at the center resonant frequency, then the impedances (admittances) of the previous slots in the waveguide just add together at the next slot. The objective is to have the total impedance (admittance) contribution of all the slots at the feedpoint be equal to the intrinsic impedance (admittance) of the waveguide. This guarantees the lowest VSWR at the center operating frequency. The slots are located along the waveguide where their current excitation is nearly maximum. Transverse currents that are in phase with the E -field excite longitudinal and side-wall slots. Therefore they are centered along the waveguide at voltage maxima. Longitudinal currents that are out of phase with the E -field excite series slots. Therefore they are centered at voltage minima along the waveguide.

Linear Resonant Slot Arrays The design of a linear waveguide slot array begins by first determining the aperture distribution, and hence the slot excitation, required to achieve the beamwidth, gain, and sidelobe level needed at the center frequency of the operating frequency band. Since the slot elements are $\lambda_g/2$ apart and in phase, the array pattern along the longitudinal plane of the waveguide is controlled by number of elements and their excitation amplitudes. Depending on the type of slots used, their amplitudes are controlled by offsetting the elements from the centerline of the waveguide, by rotating the slots at their centers, or by a combination of both offset and rotation. The square of the voltage excitation of a slot is proportional to its radiated power and its resonant conductance. For a linear array of N slot elements we have

$$\sum_{n=1}^N g_n = W \quad (9-8)$$

$$g_n = \frac{A^2(n)}{\sum_{i=1}^N A^2(i)} = KA^2(n) \quad n = 1, 2, \dots, N \quad (9-9)$$

where g_n is the resonant conductance of the n th slot normalized to the waveguide admittance Y_0 , $A(n)$ is the given aperture voltage distribution at the n th-slot location, K is the power normalization factor, and W equals 1 for end-fed arrays or 2 for center-fed arrays.

When a waveguide is fed at the center, the normalized conductances looking into both ends are in parallel at the feedpoint, which is why W is 2. Equations 9-8 and 9-9 are then solved for g_n , and the design is completed by dimensioning the slots and positioning them at the locations where they will resonate at the correct frequency with conductance g_n . Figure 9-5 shows the two preferred methods for feeding resonant arrays. For both end-fed and centered approaches, the distance between the short-circuit waveguide termination and the first end slot is $\lambda_g/4$. The short circuit reflects as an open circuit, which has an admittance of zero, to the first slot from the end. This is in parallel with the conductance of the end slot.

Linear slotted arrays utilize either longitudinal shunt slots or edge-wall shunt slots as radiating elements. Narrow longitudinal shunt slots radiate array patterns that have very high cross-polarization isolation. An example of an end-fed, dual-sided longitudinal linear slot array is an X-band array, as shown in Figure 9-6. For this array, slots have been placed on both wide walls of the waveguide. The array radiates an omni-azimuthal, horizontally polarized pattern. For a vertically polarized pattern, the array would have to be comprised of edge-wall shunt slots, which are also excited by a single transverse current. However, the cross-polarization isolation with edge slots is not as good as with the longitudinal slots because every other element is canted in a different direction, and the polarization purity is dependent on the uniform excitation of the slots and external cancellation of the cross-polarized field components. Rotated series slots also have a lower cross-polarization isolation level as compared with the longitudinal shunt slots and edge slots due to the rotation of the slots and due to some excitation of the slots by the transverse currents.

The number of slots that can be arrayed in a single waveguide section is limited. The impedance bandwidth of the waveguide array narrows rapidly with an increasing number

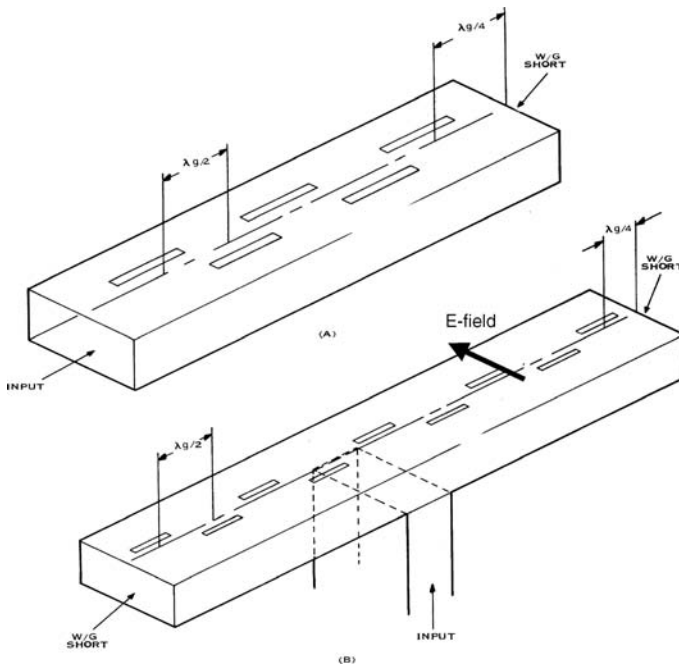


FIGURE 9-5 Linear resonant waveguide slot arrays with longitudinal slot elements. In (a) the array is fed from the waveguide end. In (b) the array is fed from the center through an E -plane T coupler. Figure 9-5b could also be fed by a series or shunt slot from another waveguide on the backside.

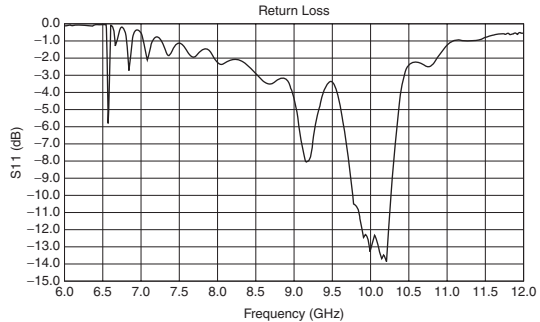


FIGURE 9-6 End-fed resonant longitudinal slot array. Slot elements are located on both sides of the waveguide to provide an almost omni-azimuthally, horizontally polarized radiation pattern at 10 GHz. The slots on either side are directly behind each other. The background can be seen through the slots. A dip of 10 dB in the radiation pattern occurs at 90°. The return loss at 10 GHz is -13 dB. (Photo: Courtesy of T. O. Perkins, BAE SYSTEMS)

of elements.¹ Any small deviation in admittance from a perfectly resonant element as was designed is rapidly transferred to other elements, causing the reactive admittance component to grow rapidly. Therefore any mismatch rapidly builds up at the feedpoint as the frequency changes. A common subarray size varies from 12 to 16 elements for an impedance bandwidth of 4 percent to 3 percent, respectively.

Planar Resonant Longitudinal Shunt and Series Slot Arrays Resonant planar arrays are used in many radar applications and preferred over reflector antennas because the beam pattern can be shaped and monopulse sum and difference ports for tracking can be implemented directly. Many present-day radars still utilize fixed beam apertures that are mechanically scanned. Others are hybrid scanned, electronically in one plane and mechanically in the other. Planar arrays are constructed by placing several linear arrays side by side, as shown in Figure 9-7. A large planar array may consist of several subarrays, and each subarray may consist of several linear arrays. Because the minimum width of a waveguide is equal to the half wavelength of its cutoff frequency, inter-element spacing between adjacent waveguides at the operating frequency is greater than $\lambda/2$, which can create grating lobes beyond certain scan angles in the plane orthogonal to the waveguides. As with standard phased array design, the radiation pattern of the array is a function of the external placement and excitation of the elements relative to each other. Therefore wall thicknesses between adjacent waveguide sections are made as thin as structurally acceptable. When the waveguide width is still too large for grating lobe-free beam scan, you can design the structure using custom waveguide of lesser width or turn the waveguide on its side and utilize edge-slot elements.

Large resonant planar arrays are usually fed using corporate or series-fed waveguide networks like the one seen on the backside of the planar array in Figure 9-8. The number of waveguide subarrays that can be fed by a secondary waveguide coupled by series or shunt slots is limited by impedance bandwidth to the same number of slot elements that can be excited in a single linear waveguide array. Coupling is done by slots rotated to an angle that

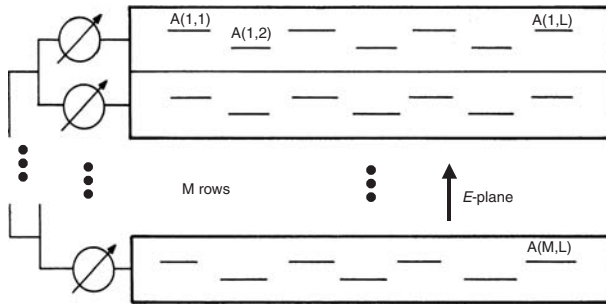


FIGURE 9-7 Resonant longitudinal shunt-slot array fed by phase shifter network located at the ends of the waveguides. This figure shows one subarray “stick” per row. Each row could have multiple sticks. This array will scan a beam in the *E*-plane, the direction orthogonal to the waveguides.

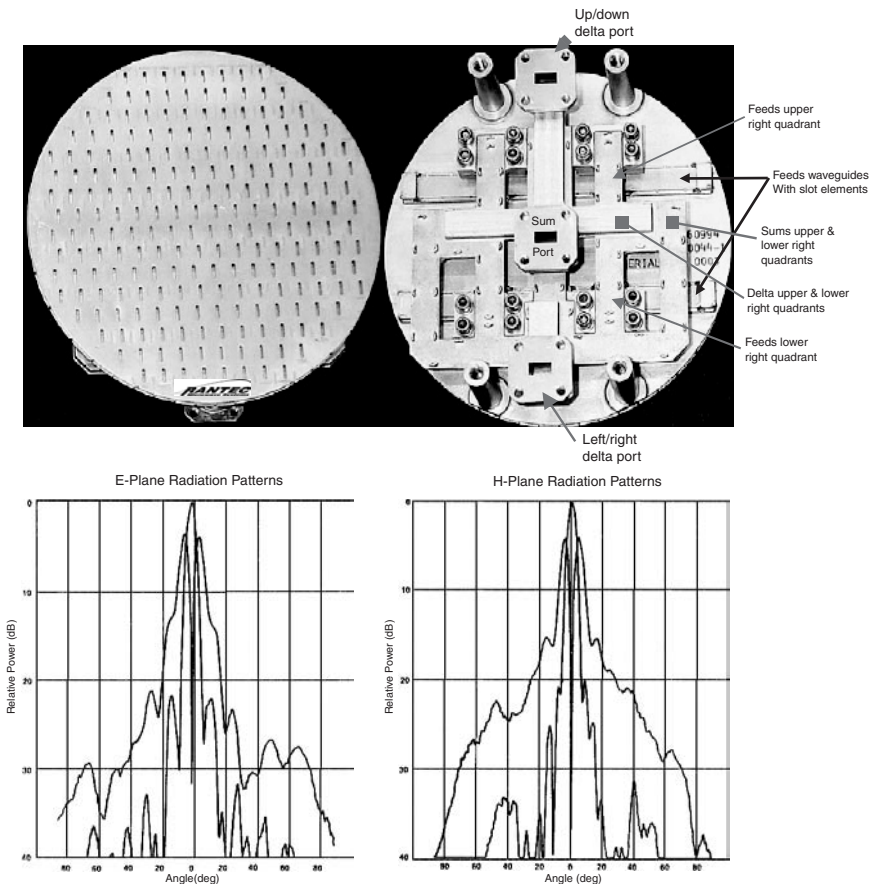


FIGURE 9-8 Ka-band resonant longitudinal shunt slot fixed beam array. The beamformer has been designed to provide up/down and left/right monopulse capability. *E*-plane and *H*-plane sum and difference patterns indicate sidelobes lower than -20 dB and a difference null depth of -32 dB.

couples the correct amount of power into each subarray. Depending on how many subarrays must be fed, several layers of waveguide might be required in the network to feed the array. In general, the series-slot center-feed system is a simpler and more compact design than the waveguide end-feed system for the same antenna size. Caution should be exercised, however, since the feed slots sometimes couple strongly to the radiating slots above them. This is especially true for series slots and edge slots. Longitudinal shunt slots do not couple strongly unless the waveguide height has been greatly reduced. For those arrays that have phase shifters, the row or column divider manifolds are brought to a common edge or plane where the phase shifter network is attached.

Series resonant slot arrays are designed in a similar fashion as longitudinal shunt-slot arrays. Because their cross-polarization isolation is not as good as the longitudinal slots, they are not often used as a primary radiator. Their use is more often as the feed distribution array behind the radiating array. The reason for their preference in this application is that the elements can remain centered about the same waveguide centerline and be able to feed other waveguides within $\lambda/2$ inter-waveguide spacing. The slot elements are positioned along the centerline of the waveguide, rotated at an angle from the centerline to couple energy from the waveguide to the slot. (The rotation angle is sometimes called the *inclination* angle.) Unlike the shunt slot, whose resonant length is affected by its offset from the centerline, the resonant length of the series slot is rather insensitive to the rotation angle. When series slot elements are placed within $\lambda_g/2$ along the waveguide, they are excited with opposite polarity. This polarity change is reversed by rotating the slot to the negative angle from the centerline. When the field radiated by each slot is separated into a component polarized along the direction of the waveguide and one orthogonal to the waveguide, the longitudinal components cancel each other and the transverse components combine. This is why the E -plane of the series slot array is also perpendicular to the axis of the waveguide. The cross-polarization problem is exacerbated when the array is fed with a tapered excitation whereby the next element has a different amplitude than the previous. In this case the external cross-polarization component of each element does not cancel itself fully, thus leaving a residual cross-polarized component.

Once the required slot excitations for either the longitudinal shunt slots or the series slots have been determined, the slot conductance or resistance is then calculated before determining the slot dimensions and offset from the center waveguide line for the shunt elements or the rotation angle for the series slots. By using the slot equivalent circuit and assuming that all the slots are at resonance, the normalized slot conductances $g_{m,n} = G_{\text{slot}}/Y_{o \text{ waveguide}}$ or series resistances $r_{m,n} = R_{\text{slot}}/Z_{o \text{ waveguide}}$ in the waveguides are determined by

$$\begin{aligned} \sum_{n=1}^{N_k} g_{m,n} &= W \\ \sum_{n=1}^{N_k} r_{m,n} &= W \end{aligned} \quad (9-10)$$

where N_k is the number of elements in a given waveguide section in row m and $W = 1$ for end feed or $W = 2$ for center feed. Let the power radiated by the slots of a waveguide subarray section be

$$B_k^2(m) = \sum_{n=1}^{N_k} A^2(m,n) \quad m = 1, 2, \dots, M \text{ and } k = 1, 2, \dots, K \quad (9-11)$$

where K is the number of waveguide sections in a given array row m . Therefore, $g_{m,n}$ and $r_{m,n}$ are given by

$$\begin{aligned} g_{m,n} &= \frac{A^2(m,n)}{B_k^2(m)} \\ r_{m,n} &= \frac{A^2(m,n)}{B_k^2(m)} \end{aligned} \quad (9-12)$$

Once g and r are known, then the design tables can be used to arrive at initial slot dimensions and positioning before attempting an optimization. Commercial software is available to assist in the design of slot arrays. It is still a numerically challenging problem to compute all mutual coupling and edge effects of moderately large arrays. Therefore, available software probably still has to make approximations to arrive at a design. When a truly high performance array is desired, such as one with very low level sidelobes or very sharp monopulse nulls, designers might have to write their own custom code to optimize the design.

Resonant Edge-Wall Shunt-Slot Array

The edge slot is cut into the narrow wall of a waveguide. The slot is designed to operate at resonance, and its length is nearly half a wavelength. Because the edge dimension of a waveguide is usually less than half its width, the depth of the slot cut runs into the broad wall. The slot length is approximately equal to its visible length on the edge wall added with twice the length of the cut depth into the broad wall, which accounts for the cuts into the upper and lower broad walls. Planar arrays of these elements are not practical because spacers between adjacent waveguides need to be placed to prevent the slots edges from being covered up. This is not the case when the whole slot can fit on the narrow edge-wall. Thus planar slot-antenna arrays often use end-loaded slots such as C or I slots as seen in Figure 9-9 to allow the slots to resonate on the edge-wall surface. There usually is no advantage in using edge-wall shunt



FIGURE 9-9 Resonant edge-wall, shunt-slot array. The slot tips are enlarged to resonate the slots at a shorter length. Note that when an edge array is fed from the center with an H -plane T or edge-wall coupler, the polarization of the elements is reversed on half of the array. (Photo courtesy of Microwave Systems, Inc.)

slots in a planar array unless closer inter-element spacing is needed to electronically scan a beam orthogonal to the waveguide axes with minimal grating lobes.

Edge-slot arrays also have difficulty achieving high cross-polarization isolation. Because the array is comprised of elements with opposing tilt angle, the cross-polarization isolation is achieved in spatial combining. If the slot amplitudes are not all of the same level, then the cross-polarization components do not fully cancel.

Edge-wall slotted waveguides are usually fed from the end because center feeds tend to couple strongly to slots on the opposite edge. The transverse currents in the waveguide flow around the walls in what could be described as a “square” perimeter orthogonal to the waveguide axis. When a slot interrupts the current flow along a wall, the currents along that whole perimeter are affected. Therefore, other shunt slots on that perimeter will couple to that current. Despite the preference for end-fed waveguides, it is often necessary in larger arrays to feed the edge-slotted waveguides from the opposite edge-wall at the center. Coupling to an edge waveguide wall is known as an *H*-plane coupler. The signals coupled to the center of the host waveguide are initially in phase before they propagate to the opposite ends of that waveguide. Therefore it is necessary to reverse the phase of the slot elements on one end of that waveguide if it is going to form a broadside beam.

Determining the excitation for the elements of the edge wall, shunt-slot elements follows the same computational EM modeling process as with the series and longitudinal slots. Once the excitations $A(m, n)$ are known, then the normalized slot conductance is computed as per Eqs. 9-10 to 9-12. The slot voltage is proportional to $V_0 \sin(\theta)$ where θ is the tilt angle from the vertical. Therefore the conductance and coupled power will be proportional to the square of the voltage. A slot design that meets the required conductance is difficult to obtain. Until recently, when results of MOM and FEM modeling of edge slots were published,^{2,3} there was no simple formulation for edge slots that cut into the broad walls. Slot characteristics were obtained through scattering parameter measurements only.

Traveling-Wave Slot-Array Design

Unlike resonant waveguide slot arrays, traveling-wave slot arrays avoid resonances or standing-wave conditions in the waveguide. Hence the slots of traveling-wave arrays are spaced by either more or less than one-half waveguide wavelength. As in the case of resonant waveguide slot arrays, resonant slots are used in traveling-wave arrays to maximize frequency-bandwidth performance. A matched load terminates the slotted waveguide to prevent the formation of a secondary beam due to reflected waves. Although much effort is put into minimizing reflections that occur from the termination and also from every slot along the waveguide, sometimes it is necessary to create controlled reflections to cancel the appearance of secondary lobes that are not due to grating lobes.

Inter-element slot spacing is selected to be close to one-half waveguide wavelength ($\lambda_g/2$). Therefore the polarity of adjacent slots is of the opposite sense. For example, adjacent longitudinal shunt slots are on opposite sides of the guide centerline, and adjacent edge-wall shunt slots are inclined to opposite sides of the vertical centerline. The alternating displacement or rotation with respect to the waveguide axis of successive slots may produce grating lobes.^{4,5} If the main beam of the traveling-wave slotted array is pointing far away from broadside, the upcoming grating lobe level may exceed the desired sidelobe level. The grating lobes can be pushed out of the visible array space by choosing a suitable inter-element spacing. The restrictions on grating lobes can be relaxed when the slot element pattern amplitude is taken into consideration at the angles where the grating lobes are appearing.

There are two types of traveling-wave slot-antenna arrays: (1) the uniformly spaced element array that produces a low-sidelobe pencil beam and (2) the nonuniformly spaced element

array that is used in shaped-beam array designs. Longitudinal shunt slots, edge-wall shunt slots, and rotational series slots are used in both traveling-wave type arrays. Traveling-wave arrays are always fed from the end of the waveguide. The progressive phase shift between elements, which changes with frequency, causes the beam to scan. This means that as power is fed into the waveguide, the first slot couples some power, leaving the remaining power to travel to the next slot, and so on. By the time the remaining power arrives at the last slot, that slot must couple most of it to meet its prescribed excitation level. Since there are too many factors that can affect the precise amount of power that travels from slot to slot, and since it is impossible for the last slot to couple 100 percent of the remaining power, the array is designed to let about 5 percent to 10 percent of the power be absorbed by the termination. This helps to stabilize the excitation of the last few slots before the termination.

The process to determine the slot conductance or resistance can be simplified as follows if the waveguide can be considered as lossless and if the reflection coefficient due to each slot is minimal. Consider the equation for the power input to the waveguide

$$P_{in} = P_{\text{all slots}} + P_{\text{load}} = K \sum_{i=1}^N p_i + P_L \quad (9-13)$$

where

P_{in} is the input power to the waveguide

$P_{\text{all slots}}$ is the power radiated by all the slots in that waveguide

$P_{\text{load}} = P_L$ is the power absorbed by the load at the end of the waveguide

p_i is the normalized power radiated by slot i given by $A^2(m,i)$, the calculated slot excitation

K is the power calibration factor given by

$$K = \frac{P_{in} - P_L}{\sum_{i=1}^N p_i} \quad (9-14)$$

If slot 1 is the slot closest to the feed end of the waveguide, then the first shunt slot conductance (resistance for series slots) is given by

$$g_1 = \frac{K p_1}{P_{in}} \quad (9-15)$$

The conductance at the second slot is found by subtracting the power absorbed by the first slot.

$$g_2 = \frac{K p_2}{P_{in} - K p_1} \quad (9-16)$$

The conductance for the third slot would include the power from the second slot in addition to the power from the first slot. This process is iterative until the last slot N . The conductance for that slot is given by

$$g_N = \frac{K p_N}{P_{in} - K \sum_{j=1}^{N-1} p_j} \quad (9-17)$$

Note that this approximate design is valid only if the largest slot resonant conductance g_{\max} satisfies the following relationship:

$$|g_{\max} \csc k_g d| < 0.447 \quad (9-18)$$

where d is the spacing between elements and $k_g = 2\pi/\lambda_g$. Through the proper choice of P_L the resultant slot conductances can usually be made to satisfy this criterion, although undesirably high power may be dissipated at the termination.

For arrays with many slots per waveguide section, the determination of the slot conductance by Eq. 9-17 is very accurate since the coupling factor has to be low. However, with fewer elements the coupling factor is higher, and the slots create VSWR mismatches. For this case, K_p determines the power that must be absorbed per element. A transmission-line circuit is drawn representing the waveguide. The slots are shown as lumped loads at their proper locations along the transmission line. In this case, the analysis begins at the end of the transmission line working up toward the feed. The assumption can be made that the load is perfectly matched unless its VSWR is known. That admittance is translated to the position of slot N . To do this, the reflection coefficient must first be determined. The lump admittance of the slot, which is not yet known, is now in parallel with the translated transmission-line admittance. Based on earlier knowledge of the power going to the load and the power going to slot N , the slot admittance can be calculated from the ratio of slot admittance to the total admittance at that point, which is the sum of the slot admittance and the translated admittance from the load. The solution should not be much different from what was calculated in Eq. 9-17. However, an effort should be made to choose a slot admittance that minimizes the reflection coefficient from the combined admittance of slot and the translated line admittance. The total admittance at slot N is now translated to slot $N - 1$ and the process is repeated. In some cases the slot admittance might need extra susceptance to minimize the reflections. There is little that can be done in slot offset and rotation to get a reactive component, but posts or other compensators within the waveguide in the vicinity of the slot might be needed.

Uniformly Spaced Traveling-Wave Slot-Array Design A uniformly spaced traveling-wave slot array can be designed to produce a low-sidelobe pencil beam with larger bandwidth than that of the resonant waveguide slot array in terms of the array radiation pattern and input VSWR. If the reflections between elements are negligible, a constant phase difference between elements results in a progressive phase shift along the array aperture. The resulting phase front points the peak of the beam to angle θ , which is determined by

$$\sin \theta = \frac{\lambda}{\lambda_g} - \frac{(N-1/2)\lambda}{d} \quad N = 0, \pm 1, \pm 2, \dots \quad (9-19)$$

where d is the inter-element spacing and $|N| > 0$ refers to grating lobes. The influence of element spacing on scan angle is as follows. Assume that d is $\lambda_g/2$ at frequency f_o . The scan angle θ is at 0° . Now assume that $d > \lambda_g/2$ for the same frequency f_o . The beam peak has now moved in angle toward the load end of the waveguide. When the spacing is chosen so that d is less than $\lambda_g/2$, the beam peak now moves in angle toward the feed end of the waveguide. However, given a fixed inter-element spacing, the beam scans toward the load end of the waveguide as the frequency is increased. When the operating frequency is close to the cutoff frequency, grating lobes begin to appear in the radiation pattern of the array. Filling the waveguide with a dielectric material extends the frequency range, and therefore the angle, over which a grating lobe-free beam scan is possible. Hence, the ability to adjust the beam peak to a position around which a frequency scan can occur is important for traveling-wave arrays.

The nonresonant array achieves its large impedance bandwidth by virtue of the phase differences between the reflections from the various slots. The phase differences, which arise from nonresonant spacing, cause the resultant sum of all the reflected waves to be quite small. When the slot spacing is approaching $\lambda_g/2$ and the beam peak is very close to the broadside position, the input VSWR will increase rapidly as the slot admittances add in phase so that the input-impedance magnitude is much larger than unity. For this reason a broadside beam direction is reserved only for resonant slot arrays.

Nonuniform Traveling-Wave Slot-Array Design Some special applications, such as in air traffic control, need the shaped beam to scan over a sector. The aperture illumination of a shaped beam requires, in general, both amplitude and phase variations across the aperture. Therefore it is desirable to know whether and how a linear slot array can produce a shaped beam.

It will be seen in the next section that the aperture illumination amplitude can be controlled by the slot conductances. This is the same as was done with the resonant slot arrays. Varying the slot location along the waveguide and, to some extent, varying slot reactance, can control phase. The latter provides limited phase control and results in a narrow frequency bandwidth. A better method for varying the phase is the use of resonant slots and nonuniform inter-element spacing. The resonant slot conductance or resistance controls the aperture illumination amplitude, while the nonuniform element spacing provides the required phase distribution across the aperture.

The process to determine the locations of individual slots along a waveguide to create a desired aperture excitation is an iterative one. The first iteration is to design an array with uniform spacing where the desired phases at each element are specified. A particular waveguide section will feed a subarray of the whole array. If the phases of the slots that would be part of that waveguide section are all normalized to the slot closest to the load termination, then relative phases for each slot up the waveguide section are known. The next step is to compare those phases to the slot phase excitation provided by the waveguide. The difference between the desired phase and the waveguide excitation is how much phase perturbation is required.

Assume that the synthesized aperture illumination for an array is given by

$$A(d_i) = |A(d_i)| e^{-j\psi(d_i)} \quad (9-20)$$

where A and ψ are the amplitude and phase excitations required at slot d_i . This is a complex version of the slot excitation $A(m, n)$ used in earlier discussions except with the ability to move the slot position. For simplicity, waveguide losses and reflections from the slots are assumed to be negligible. Under these assumptions, the waveguide excitation phase is given by

$$\psi_g(d_i) = k_g d_i + (N - i)\pi \quad (9-21)$$

where i signifies the i th slot, N is the total number of slots in that waveguide section, d_i is the distance from the last slot N to slot i , k_g is the waveguide wave number, and ψ_g is the phase excitation coupled from the waveguide. The phase error is given by

$$\Delta\psi_i = \psi(d_i) - \psi_g(d_i) \quad (9-22)$$

The delta phase error should be rewritten as a $\text{Mod}(\Delta\psi, 2\pi)$. Assuming that $\Delta\psi$ is not large, then it can be represented as

$$\Delta\psi_i = k_g \Delta d_i \quad (9-23)$$

A positive Δd_i represents movement toward the feed. A negative Δd_i represents movement toward the termination end of the waveguide. Once the slots are repositioned, the new slot positions are submitted to the aperture design model to reassess the mutual coupling effects and to see if the new results agree with the modified slot excitations at their new positions. If the phase differences are still too large, the slot repositioning process is repeated. Usually the rate of convergence is quite rapid (two or three times) if the desired illumination phase function is a slowly varying function, such as the phase function for a cosecant-squared-beam aperture. After the slot locations have been determined, the slot resonant conductances can be computed in the same manner as for the uniform traveling-wave array.

If the waveguide losses or the reflections from the slots are not negligible, then the transmission-line model of the slotted waveguide is required to determine the phase of the slot coupling to the waveguide. The simple solutions obtained by neglecting the reflection from the slots and waveguide losses are employed as the first-order approximation in the iteration procedure. The iterative procedure is the same as has been described earlier, but ψ_g and $\Delta\psi$ must be expressed by the transmission-line model to compute the phases associated with the admittance and the slot resonant conductances. The geometry of the slots can then be realized after the slot locations and resonant conductances have been determined.

In general, if the nominal value of the unequal element spacing is larger than one-half free-space wavelength, higher sidelobes will appear in the region far off broadside. Therefore, it is recommended that nominal inter-element spacing be less than one-half waveguide wavelength for the nonuniformly spaced slotted array.

9.4 COMPUTATION METHODS FOR SLOT CHARACTERIZATION

Before the advent of more capable EM modeling tools, extensive analytical and experimental studies were conducted to characterize waveguide-fed slots. Automated test equipment, such as computerized network analyzers, has greatly simplified data recording and processing to enable full S-parameter measurement techniques⁶ to characterize the effects of slot geometry and placement parameters. On the other hand, the cost of conducting elaborate empirical studies is prohibitive, and many recent investigations have focused on improving the computation models of specialized parts such as waveguide *T*-junctions, waveguide-to-waveguide slot coupling, and finite array performance predictions using slot radiators. Few recent publications attempt to develop explicit formulas that predict waveguide-fed slot admittance in the presence of a finite array because these parameters are obtained directly through EM modeling. Commercial software provides design information for waveguide arrays. However, much work still needs to be done to develop modeling tools that can provide an optimized design solution. Still, many waveguide array designers utilize existing formulas for slot admittance to obtain a first order design that might not include all the mutual coupling effects, edge effects, radome scattering effects, etc. Then they utilize numerical optimization techniques to complete the design. This section will focus on the basic waveguide-fed slot conductance (admittance) representations published in earlier works.

Stevenson's Solution

Stevenson⁷ developed his formulas for representing slot characteristics by making the following assumptions: the slot was cut in a perfectly conducting, infinitely thin wall; the width of the slot was much less than its length; the slot length was nearly equal to $\lambda_0/2$; and the slot was radiating over a perfectly conducting ground plane of infinite extent.

Using transmission-line theory and the waveguide modal Green's functions, Stevenson derived the values of the resonant resistance and conductance, normalized to the waveguide impedance, for various slot types along a rectangular waveguide.

Longitudinal Shunt Slots: The conductance of a longitudinal shunt slot (slot a in Figure 9-1) cut in the broad face of a waveguide is given by

$$g = g_1 \sin^2 \left(\frac{u\pi}{a} \right) \quad (9-24)$$

$$g_1 = \frac{2.09a\lambda_g}{b\lambda} \cos^2 \left(\frac{\lambda\pi}{2\lambda_g} \right) \quad (9-25)$$

where λ is the free-space wavelength, λ_g is the guide wavelength, u is the slot displacement from the waveguide centerline, and a and b are the waveguide width and height.

Series Slots: The resistance of the centered inclined/rotated series slot (slot d of Figure 9-1) in the broad face of a waveguide is given by

$$r = \frac{0.131\lambda^3}{ab\lambda_g} [I(\phi) \sin \phi + \frac{\lambda_g}{2a} J(\phi) \cos \phi]^2 \quad (9-26)$$

$$\frac{I(\phi)}{J(\phi)} = \frac{\cos(\pi\xi/2)}{1-\xi^2} \mp \frac{\cos(\pi\zeta/2)}{1-\zeta^2} \quad (9-27)$$

$$\begin{cases} \xi \\ \zeta \end{cases} = \frac{\lambda}{\lambda_g} \cos \phi \pm \frac{\lambda}{2a} \sin \phi \quad (9-28)$$

where ϕ is the slot inclination angle with respect to the centerline of the waveguide.

Narrow-Wall (Edge) Shunt Slots: The conductance of an edge slot on the narrow face (slot c of Figure 9-1) is given by

$$g = \frac{30\lambda^3\lambda_g}{73\pi a^3 b} \left[\frac{\sin \phi \cos \left(\frac{\pi\lambda \sin \phi}{2\lambda_g} \right)}{1 - \left(\frac{\lambda \sin \phi}{\lambda_g} \right)^2} \right]^2 \quad (9-29)$$

where ϕ is the included angle of the slot from vertical.

These equations yield quite accurate results for a single-slot radiator without mutual coupling being included and give no information about the reactive component. However, mutual-coupling effects are not negligible especially with $M \times N$ arrays, and Stevenson's approximations could be off by a large margin.

Variational Technique

Oliner⁸ first formulated an equivalent circuit model for a thick slot using a variational technique. An accurate impedance solution can be obtained by using a slot-aperture field

close to the actual slot field. The method can also account for mutual coupling and provide an estimate for the reactive component of the slot impedance.

Consider a longitudinal shunt slot in the broad face of a rectangular waveguide radiating into free space. By extending Oliner's variational formulation, an explicit impedance expression can be derived in terms of the slot-aperture admittance $Y_r = G_r + jB_r$.⁹

Let the length and the width of the rectangular slot be a' and b' , respectively. The propagation constant and the characteristic admittance of the TE_{10} mode in the waveguide are given by $\kappa = (k^2 - (\pi/a)^2)^{1/2}$ and $Y_0 = \kappa/\omega\mu$, where $k = 2\pi/\lambda$, $\omega = 2\pi f$ is the angular frequency, and μ is the permeability. The rectangular slot on the thick waveguide wall is regarded as a stub waveguide with propagation constant $\kappa' = (k^2 - (\pi/a')^2)^{1/2}$ and characteristic admittance Y_0' . The input admittance at the stub-guide junction can be determined by

$$Y_{rj} = Y_0' \frac{Y_r + jY_0' \tan \kappa' t}{Y_0' + jY_r \tan \kappa' t} \quad (9-30)$$

where t is the thickness or depth of the slot. From the standpoint of the feed waveguide, the equivalent circuit of a displaced longitudinal shunt slot can be represented by a shunt network, as depicted in Figure 9-10. Z_{rj} is the impedance due to the radiating aperture, while X_j is related to the stored power in the feed waveguide. Near resonance, these quantities can be approximated by the variational result.⁹

$$\frac{R + jX}{Z_0} = \frac{Z_{rj} + jX_j}{Z_0} = \frac{a'b'}{2N_s^2} \frac{(Y_{rj} + jB_j)}{Y_0} \quad (9-31)$$

$$N_s^2 = (2/ab) \frac{2\pi^2 b'}{aa'\kappa} \left[\frac{\sin(\pi u/a) \cos(\kappa a'/2)}{(\pi/a')^2 - \kappa^2} \right]^2 \quad (9-32)$$

where u is the displacement of the slot from the center of the waveguide. The junction susceptance B_j is given by

$$B_j = \frac{4b'}{\omega\mu ab} \sum_{m=0, n=0} \left[\cos\left(\frac{m\pi}{2} + \frac{m\pi u}{a}\right) \text{sinc}\left(\frac{m\pi b'}{2a}\right) \right]^2 \frac{\text{Im}(W)}{(\epsilon_m \epsilon_n)} \quad (9-33)$$

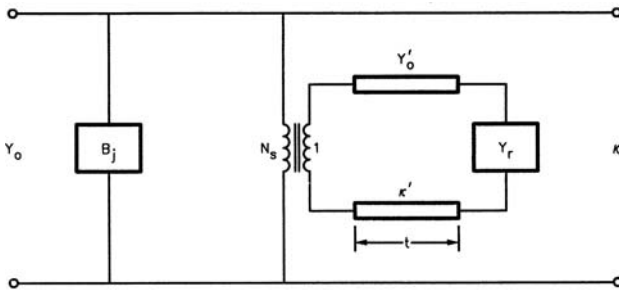


FIGURE 9-10 Equivalent network of a longitudinal shunt slot

where

$$\text{sinc}(x) = \frac{\sin(x)}{x}$$

and

$$Q_2(m, n) = \frac{\kappa'^2}{\left[\left(\frac{\pi}{a} \right)^2 - \kappa_{mn}^2 \right]} \quad (9-34)$$

$$Q_3(m, n) = 2 \left(\frac{\pi}{a'} \right)^2 \left[\frac{\left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2}{\left[\left(\frac{\pi}{a'} \right)^2 - \kappa_{mn}^2 \right]^2} \right] \left[\frac{1 + e^{-ja'\kappa_{mn}}}{ja'\kappa_{mn}} \right] \quad (9-35)$$

$$\kappa_{mn} = \left[k^2 - \left(\frac{m\pi}{a} \right)^2 - \left(\frac{n\pi}{b} \right)^2 \right]^{1/2} \quad (9-36)$$

$$\epsilon_n = \begin{cases} 2 & n = 0 \\ 1 & n > 0 \end{cases} \quad (9-37)$$

Im indicates the imaginary part of the complex expression.

Resonance can be defined as the condition in which slot admittance is a real quantity. To compute the slot resonant length, several slot admittances corresponding to several values of a' nearly equal to $\lambda_o/2$ are computed first. The slot resonant length can then be determined by interpolating the length that yields zero imaginary part for the admittance. The preceding solution applies to rectangular slots only. For round-ended slots, the resonant length must be modified by

$$L = a' + Cb'(1 - \pi/4) \quad (9-38)$$

where a' is the resonant length of the equivalent rectangular slot and C is the correction factor.⁹

A similar variational formulation has been applied to compute the impedance of a rotated series slot on the broad side of a rectangular waveguide. The slot rotation angle replaces the slot displacement variable u from the waveguide centerline in the derivation. However, the slow convergent series in the resultant reactance expression requires special numerical treatment. Because field distributions in complicated slot shapes are difficult to obtain, variational solutions for edge slots and folded slots have not been derived explicitly.

Scattering Modal Analysis

Using an approach similar to Frank and Chu's formulation of the E -plane T solution,^{10,11} Montgomery derived the slot-characterization scattering matrices and hence the shunt-slot admittance and series-slot impedance.¹² This scattering modal analysis technique requires a greater computation effort than Stevenson's and the variational solutions.

The slot is assumed to be a stub guide branched out from the main guide where the field in the slot is expressed by a series of stub-guide modal functions. The main-guide field is represented by a series of modal Fourier integrals in which the wave number corresponding to the guide axis is the integration variable. By matching the tangential slot and waveguide field components at the slot aperture, two integral equations are obtained. The expansion coefficients and the scattering parameters can then be determined by solving these two integral equations using the Ritz-Galerkin method. After the expansion coefficients have been computed, the junction scattering matrix of the slot in the waveguide can be calculated. By combining the junction scattering matrix with the reflection coefficient from the radiating slot, the slot equivalent circuit is determined.

This method has been applied to the computation of individual slot characteristics and to the analysis of coupling effects between two slots or between a slot and a waveguide junction. Numerical results of longitudinal-slot parameters computed by the variational method and the scattering modal analysis compare favorably.

Moment Method

The method of moments (MOM) has been applied to characterize a narrow slot in the broad wall of a rectangular waveguide^{13,14} and can be extended to obtain the solution of folded slots in the edge wall. The moment method requires greater computational effort than any of the three other methods discussed here.

Vu Khac and Carson¹³ formulated the slot characterization by shorting out the slot and replacing it with an equivalent magnetic current. The appropriate magnetic Green's functions in the feed guide, inside the slot cavity (taking into account the finite wall thickness), and in free space are employed to derive the basic integral equations. The slot field is expressed as a series of discrete impulse functions. By matching the tangential components at the slot aperture and using the Ritz-Galerkin method, a matrix equation is derived. Then the expansion coefficients are determined by solving the matrix. The computational effort can be greatly reduced by using full-wave basis functions as shown by Lyon and Sangster.¹⁴ The computed results usually closely match the variational solution and experimental results.

When the moment method is applied to characterize the rotated series slots, a slowly convergent series is encountered, as with the variational formulation and the scattering modal analysis. All the matrix elements used in the moment method involve one or two slowly convergent series compared with only one slowly convergent series in the variational technique; therefore, the moment method requires much more computation time.

Method-of-moments (MOM) analysis of slotted-waveguide antennas has progressed rapidly during the past 20 years. Applications include the characterization of broad-wall longitudinal slots, edge-wall inclined slots,¹⁵ *I*-slots,¹⁶ and ridge-waveguide-fed slots.¹⁷ Single slot characterizations are easily extended to finite-array solutions. During the last ten years new numerical techniques in low-frequency methods have been developed to solve large impedance arrays. That makes it possible to model moderately sized arrays exactly.

9.5 DESIGN PARAMETERS

The measured and computed slot characteristics presented in this section, except for the admittance shown in Figure 9-17, are given for a 1.5875-mm wide slot in a standard WR-90 X-band waveguide at 9.375 GHz. This slot will subsequently be referred to as

the *baseline slot*. A planar slotted-waveguide array could be designed by using data presented in this section.

Longitudinal Shunt Slots

Watson,¹⁸ Stegen,¹⁹ and Kaminow and Stegen²⁰ performed extensive empirical studies of isolated longitudinal shunt slots radiating into free space. From the measured data, Stegen found that the ratio of conductance to resonant conductance G/G_m and the ratio of susceptance to resonant conductance B/G_m versus the ratio of slot length to resonant length are independent of the slot displacement off the waveguide centerline. These universal curves of the baseline-slot admittance normalized by the resonant conductance versus slot length normalized by the resonant length are plotted in Figure 9-11. The measured ratio of slot resonant length to wavelength and the measured ratio of slot resonant conductance to waveguide wave admittance versus the slot displacement off the waveguide centerline are shown in Figure 9-12 and Figure 9-13, respectively.

By using the variational result of Eqs. 9-30 through 9-33, the computed slot admittance normalized by the resonant conductance is also shown in Figure 9-11. The aperture admittance of a slot radiating into the half space is approximated by the solution given by Kurss and presented in Oliner.⁸ However, although the computed results agree well with the measured data, the ratios G/G_m and B/G_m versus the normalized slot length L/L_0 are not exactly independent of the slot displacement off the guide centerline.

By using the same technique, the computed slot resonant length and resonant conductance are given in Figure 9-12 and Figure 9-13, respectively, for comparison with the measured results. Note that the round-ended-slot resonant length L_0 is approximated by Eq. 9-38. Choosing C to be 0.64 and comparing the computed and measured results show that the computation method is sufficiently accurate for practical applications.

In airborne radar applications, large arrays of uniformly spaced slots are commonly used. The longitudinal-shunt-slot admittance of a slot in the central region of a large uniform array can be computed by the variational technique.⁹ If the element spacing of a square grid is assumed to be equal to $0.71\lambda_0$, the slot-aperture admittance can be determined by an infinite array of open-ended rectangular waveguides.²¹ The computed normalized

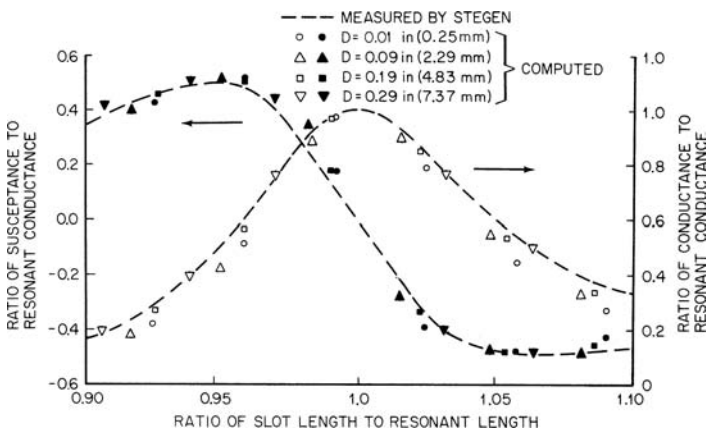


FIGURE 9-11 Admittance of the baseline longitudinal shunt slot versus normalized slot length

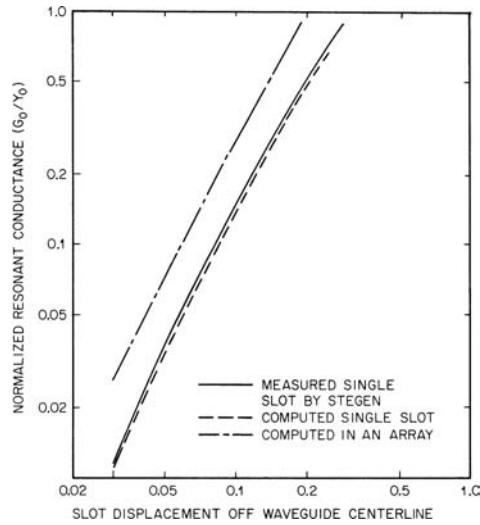


FIGURE 9-12 Normalized resonant conductance of the baseline longitudinal shunt slot versus displacement off the waveguide centerline

admittance of the baseline slot in the array is compared with the measured results of the isolated slot in Figure 9-14. The comparison shows that the mutual-coupling effects reduce the slot bandwidth performance.

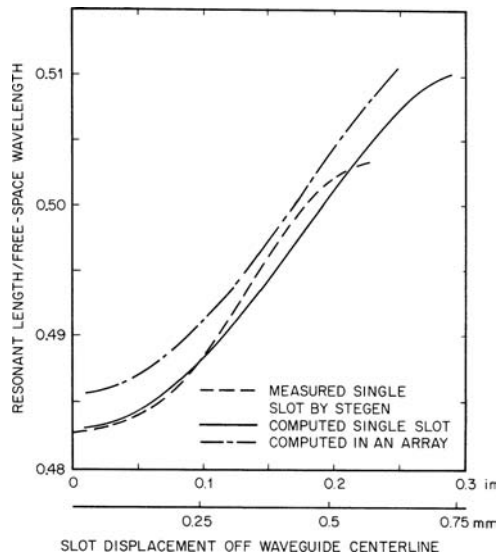


FIGURE 9-13 Normalized resonant length of the baseline longitudinal shunt slot versus displacement off the waveguide centerline

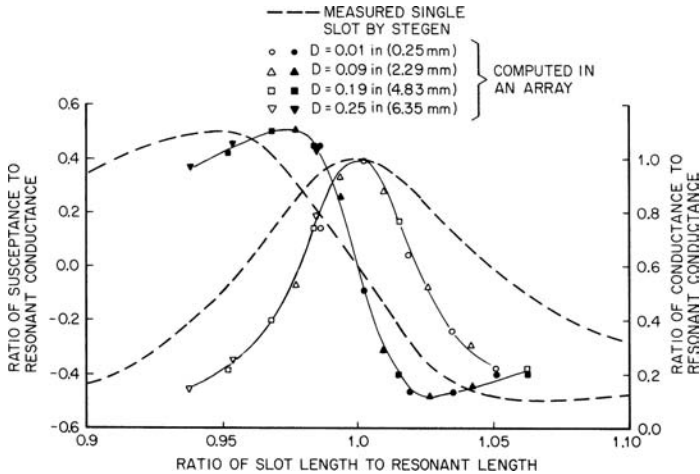


FIGURE 9-14 Admittance of the baseline longitudinal shunt slot versus normalized resonant slot length. The solid curves are best fit to the computed points. The mutual coupling effects of the array environment narrow the admittance bandwidth of the slot.

By using the same dimensions, the resonant conductance and resonant length of a slot in a large array are shown in Figure 9-12 and Figure 9-13. The mutual coupling effects also affect the slot resonant length as well as the slot resonant conductance. The small change in the slot resonant length in this case may not significantly affect antenna performance. However, an increase of the slot resonant conductance by a factor of 2 usually causes a 2:1 voltage-standing-wave-ratio (VSWR) at the center frequency. This has been observed in slot-antenna arrays designed by using measured single-longitudinal-slot conductances.

Stern and Elliott²² investigated the resonant characteristics of broad-wall longitudinal slots using the MOM formulation. They found that the single-shunt lump-circuit representation for a longitudinal slot is valid only if the slot offset from the waveguide centerline is not too large and that the waveguide height is greater than $\lambda_0/10$.

Based on the sinusoidal representation of the slot aperture field near the first resonance, Elliott²³ improved the small-array design technique to include dielectric-filled slots. An array of magnetic currents derived from the sinusoidal-slot aperture fields replaces the dipole-array solution in the original formulation to account for the external mutual-coupling effects. The internal field formulation is modified accordingly.

The MOM formulation of a finite array of longitudinal slots yields accurate results in comparison with measured data. The single lumped admittance representation of the slots in the MOM solution verifies that the infinite-array solutions shown in Figure 9-12 through Figure 9-14 are accurate for a finite array as small as 25 elements. Figure 9-15 shows the comparison between longitudinal-slot resonant conductances and the resonant lengths computed by the infinite-array approximation and the finite-array MOM results. The nominal array element spacing is $0.71\lambda_0$ in a square lattice. The finite-array results apply to uniform square arrays of N by N elements. Thus infinite-array admittances yield accurate values for large slotted-waveguide arrays.

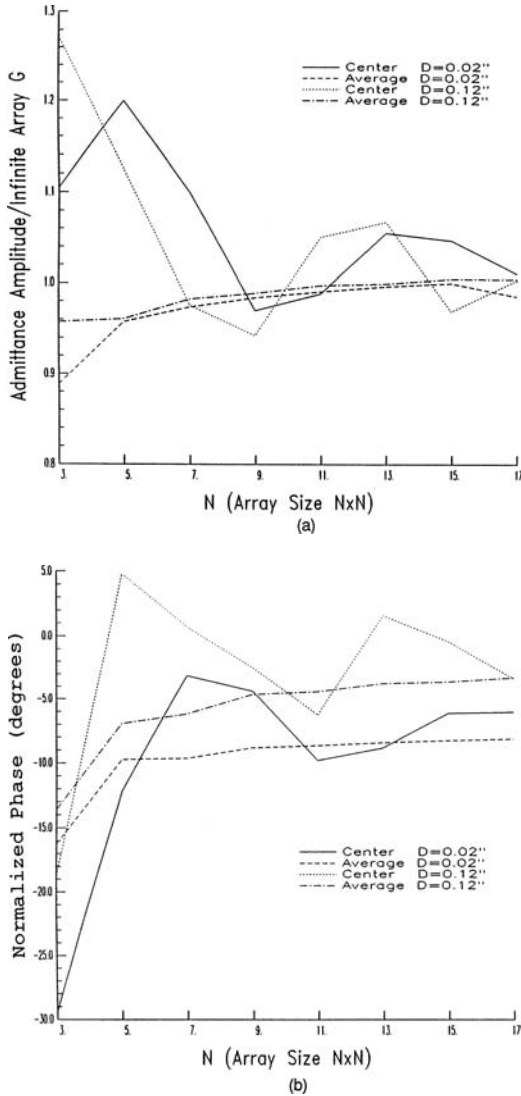


FIGURE 9-15 Slot admittances from a finite array of $N \times N$ elements normalized by infinite array results: (a) admittance amplitude and (b) admittance phase

Edge-Wall Slots

Edge-wall slots are cut into the narrow wall and wrapped around the broad wall of the rectangular waveguide. Folded C - and I -slots are more convenient for edge-wall planar arrays.^{2,3,16,24,25} Recently, several authors derived analytical solutions for narrow-wall slot admittance. These solutions include the inclined-edge slot, C -slot, and I -slot. These slot

types radiate relatively high cross-polarized fields as compared with the broad-wall longitudinal slots. Experimental results must be used to characterize them. Fortunately, strong mutual coupling effects can be measured by a waveguide section with a sufficient number of identical edge-wall slots. Watson¹⁸ has defined the incremental conductance caused by the mutual-coupling effect as the increase in the conductance of a group of resonant, half-wave-spaced, identical edge-wall slots.

For both the linear and *C* edge slots, the resonant conductance may be approximated by

$$G = G_0 \sin^2 \phi \quad |\phi| \leq 15^\circ \quad (9-39)$$

where ϕ is the inclination angle from the vertical. The slot resonant conductance is insensitive to variations in the slot resonant length. Hence, if the incremental conductance is obtained by experiment for one value of angle, Eq. 9-39 can be used to obtain the resonant conductance for other slot inclination angles with reasonable accuracy.

Plotted in Figure 9-16 are the single-slot resonant conductance and the incremental conductance for edge slots cut on the standard WR-90 rectangular waveguide as was measured in Watson.¹⁸ The incremental conductance data are sufficiently accurate for slots in a linear array far removed from the ends. The conductance of the slots near the ends of the array lies between that of the single-slot data and that of the incremental conductance.

The edge-slot admittance, as a function of the depth the slot cut has penetrated into the rectangular-waveguide broad faces, was investigated by Watson.¹⁸ Measurements were performed on a slot that was cut into standard WR-284 waveguide. The admittance for a single 15° slot with a width of 6.35 mm at 2.8 GHz is given in Figure 9-17. These results can be translated to the frequency performance of the edge slot with the conclusion that the frequency bandwidth of edge slots is greater than that of longitudinal shunt slots.

To account for the external field in the inclined-slot analysis, Jan et al¹⁵ modified the 90° wedge Green's function to approximate the thick-edge-wall Green's function. They used a MOM formulation to find the admittance of the inclined narrow-wall slots. The point-matching-technique formulation for the waveguide field usually results in a divergent

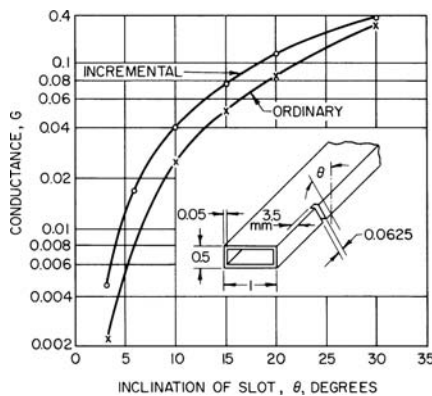


FIGURE 9-16 Incremental and single-slot conductance (unmarked dimensions of waveguide section are in inches); frequency = 9.375 GHz, WR-90 waveguide

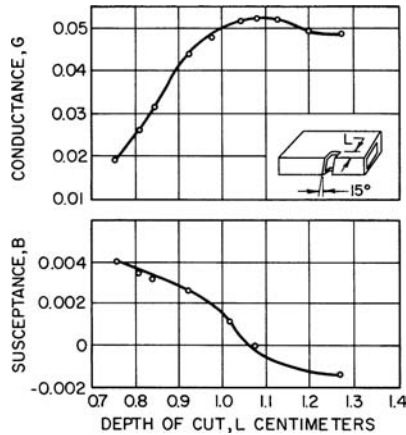


FIGURE 9-17 Variation of slot admittance with depth of cut; frequency = 2.8 GHz, WR-284 waveguide

series for inclined slots. Nevertheless, their computed results of single slots agreed with the experimental data shown in Figure 9-16 and Figure 9-17 and followed the same trend as computed by Eq. 9-39.

Employing the Galerkin method in the MOM formulation, Yee and Stelitano¹⁶ obtained a MOM solution for symmetric edge-wall *I*-slots on an infinite ground plane. They analyzed each *I*-slot as three rectilinear slots tied together and used sinusoidal basis functions in each rectilinear slot. In contrast to adding a set of junction basis functions, it took advantage of the natural solution to determine the field in the overlapped regions. The computed and measured results show that (1) the *I*-slot can be represented by a shunt admittance with additional shunt capacitance; (2) the cross-polarization isolation of an *I*-slot is about the same as other edge-wall slots (inclined slot, *C*-slot, etc.); and (3) the *I*-slot can radiate a sufficiently large dynamic range of power for practical array designs by changing its cross-branch offset.

Series Slots

The series-slot impedance can be computed by the variational formulation,⁸ the scattering modal analysis technique,¹² or the moment method¹³ as described earlier. By using either the variational formulation or the scattering modal analysis, the slot is first considered as a branched waveguide radiating into the main waveguide. The aperture impedance of the branched waveguide can be determined by assuming a cosine distribution across the slot aperture. The slot-aperture impedance facing the other waveguide is computed in the same manner except that the slot inclination angle is the complementary angle. Properly combining these two solutions yields the series-slot impedance and hence the slot resonant resistance and resonant length. When round-ended series slots are used in the design, the slot resonant length can be determined by setting $C = 1$ in Eq. 9-38.

By using the scattering modal analysis solution, the computed resonant resistance and round-ended resonant length versus the inclination angle for the baseline slot are plotted in Figure 9-18 and Figure 9-19. The input and output waveguides are assumed to be perpendicular to each other for these computed data.

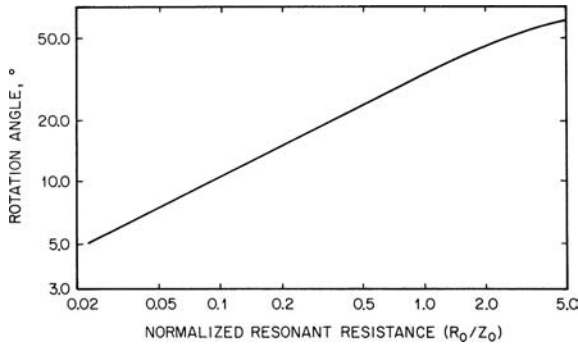


FIGURE 9-18 Normalized resonant resistance versus rotation angle of a baseline inclined series slot

Other Slot Admittance Determination Approaches

Active Slot Admittance Elliott^{26,27} derives two expressions for slot-array active admittances in terms of slot excitation voltages, slot parameters, and the complementary dipole array active impedances. By combining these two expressions with the corresponding dipole array solution, a small-array slot geometry can be specified by using the iterative procedure as described by Elliott.²⁶ The single-slot admittances of a single-slot radiator are used to model the slot admittance characteristics to reduce computational effort.

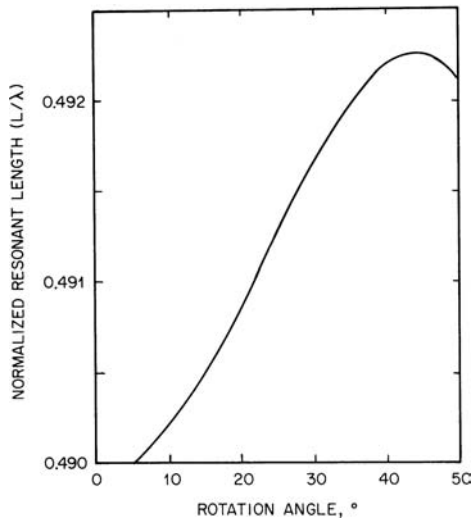


FIGURE 9-19 Normalized resonant length versus rotation angle of a baseline inclined series slot

This technique is quite accurate for small-array design as demonstrated by Elliott.²⁶ The required computation effort makes this technique unattractive for large-array applications.

Equivalent Self-Admittance Slot equivalent self-admittances can be computed by Eqs. 9-30 through 9-33 by neglecting the edge effects, the small alternating displacements from the guide centerline, the slowly varying illumination, and the small slot-length variation. The slot admittance is determined by either (1) using the equivalent infinite array slot admittance²¹ or (2) by applying Babinet's principle that relates slot aperture admittance to the complementary dipole radiation impedance.²⁸ By using the computed admittance values corresponding to several slot lengths, the resonant conductance and resonant length versus slot displacement similar to that shown in Figures 9-12 and 9-13 can be determined. By combining these design curves with the predetermined slot locations and resonant conductances, the slot-array antenna design is completed.

The computed slot admittances take mutual-coupling effects into account, but neglect edge effects, slot displacements from the guide centerline, nonuniform aperture illumination, and slot-length variation. For large arrays in general, displacements are small, illumination is slowly varying, and edge effects can be neglected. This technique is much more efficient than the active slot admittance method. For a small array, edge effects can be taken into account by several sets of resonant-conductance and resonant-length curves by using the corresponding small-dipole-array solutions. Dipole arrays corresponding to several uniformly alternating displacements can be used to compute the design curves for central elements and edge elements. For practical applications, neglecting alternating displacements in the equivalent dipole array computation results in negligible conductance error and in less than 1.2 percent resonant-length error.

A dielectric cover on the outer waveguide surface is frequently used to protect the antenna from the environment. In general, dielectric loading has little effect on the resonant conductance or resistance of a slot. However, the resonant length is greatly reduced by a dielectric cover and depends on the dielectric constant and thickness. Dielectric-loading effects can be taken into account by using the infinite waveguide array solution.²¹

Incremental Conductance The two methods described earlier are applicable only to longitudinal shunt slot arrays. Analytical solutions have not been as developed for edge-wall slots, and therefore the admittance for a particular slot design must rely on experimentation unless a more elaborate MOM or FEM model is developed.

Many waveguide test sections are required to characterize the slots, including mutual-coupling effects. Each of the test arrays has a sufficiently large number of identical conductance resonant slots so that the edge effects can be neglected. The moving lossy short technique has been successfully applied to the slot-characteristic measurements.^{6,29} Design curves such as resonant conductance and resonant length versus inclination angle for edge slots can be determined as shown in Figure 9-16. Observe that this technique is adequate for large linear-array designs but that inaccurate aperture illumination will be expected when it is applied to small-array designs.

Semiempirical Admittance The three methods discussed earlier are applicable to large arrays or arrays of equally spaced edge-wall shunt slots. Hence the semiempirical admittance technique is generalized to smaller arrays. Basically, the dipole array solution is combined with the following equations to obtain the slot characteristics semiempirically. From Eqs. 9-31 and 9-38, the slot resonant conductance can be expressed by

$$g = U' / \tilde{r} \quad (9-40)$$

where \tilde{r} is the complementary dipole active resistance normalized by the waveguide characteristic impedance and

$$U' = U(a, b, a', b') \begin{cases} \sin^2\left(\frac{\pi D}{a'}\right) & \text{Longitudinal shunt slots} \\ \sin^2 \phi & \text{Edge-wall shunt slots} \end{cases} \quad (9-41)$$

In general, the change in slot resonant length due to mutual coupling is small, and U is a slowly varying function of a' . Thus, U can be determined approximately by measurement. By combining the approximate value of U , the complementary dipole array solution, the resonant slot conductance, and the slot-location design procedure to be described later, the slot geometry can be determined in the following steps:

Step 1 Measure the single-slot characteristics.

Step 2 By substituting the complementary single-dipole resistance (zero-order approximate value of r) and the measured single-slot resonant conductance into Eqs. 9-40 and 9-41, the values of U for a given slot length and width are determined.

Step 3 By combining the single-slot solution with the equations for specifying the slot locations and resonant conductances such as Eqs. 9-8, 9-9, and 9-21, the first-order design of the slot locations and displacements (or inclination angles) can be determined.

Step 4 The driving-point impedance of the complementary dipole array can be computed by the technique of Carter,²⁸ using the computed slot locations and displacements (or inclination angles) obtained from Step 3 or Step 5 for the lower-order approximation.

Step 5 By substituting the dipole array results from Step 4 into Eqs. 9-40 and 9-41 and combining with slot-location and conductance equations, the next-higher-order approximate solutions for slot locations and displacement (or inclination angles) can be obtained.

Step 6 Repeat Steps 4 and 5 until a satisfactory convergence solution is achieved.

Note that this design procedure yields no information on the resonant lengths of the slots in the array. Fortunately, the slot admittance phase variation introduced by mutual coupling is, in general, small compared with π . Suppose that the single-dipole driving-point impedance is represented by $Z'_d = R'_d + jX'_d$ and the final solution is $Z_d = R_d + jX_d$, then if the guide-wall thickness is negligibly small, the phase difference between the single-slot admittance and the admittance of the slot in the array ψ can be approximated by

$$\tan \psi = 2(X'_d - X_d) / (R'_d + R_d) \quad (9-42)$$

The phase correction can be determined by experimentation or computation of the complementary dipole array by using the ratio of the incremental phase to the incremental slot length near the first resonance. This quantity is almost a constant and depends only on the slot and waveguide geometry.

The preceding iterative procedure has been found to converge rapidly when applied to a planar array of nonuniformly spaced C-shaped slots.

9.6 POWER-HANDLING CAPABILITIES OF SLOTTED-WAVEGUIDE ARRAYS

Slotted-waveguide arrays are required to operate at high-power levels in many applications. In some cases the slot-antenna arrays can handle the high power without undue difficulty at sea-level atmospheric pressure, but they may fail to withstand the power at high altitudes, where atmospheric pressure is much lower. When high-power capacity is required for an airborne antenna, the power-handling capability at the operational altitude must be engineered into the designs.

The power-handling capacity of a slotted-waveguide array depends on both waveguide manifold design and slot design. Following the procedure established by Gilden and Gould,³⁰ the waveguide manifold power-handling capability can be estimated with reasonable accuracy. As an example, consider a standard X-band WR-90 waveguide operating at a frequency of 9.375 GHz. On the basis of Gould's data, the continuous-wave power and the power of a 1- μ s rectangular pulse versus the altitude are shown in Figure 9-20. Observe that the maximum power level that can be handled by the waveguide decreases rapidly when the altitude increases and the waveguide pressure follows atmospheric pressure at altitude. Also note that these results apply to room temperature for a perfectly matched waveguide section. If this is not the case, the pressure at other temperature ρ is required for computing the maximum power level where

$$\rho = \rho_0(293/T) \quad (9-43)$$

T is the absolute temperature ($^{\circ}\text{K}$), and ρ_0 is the pressure at room temperature.

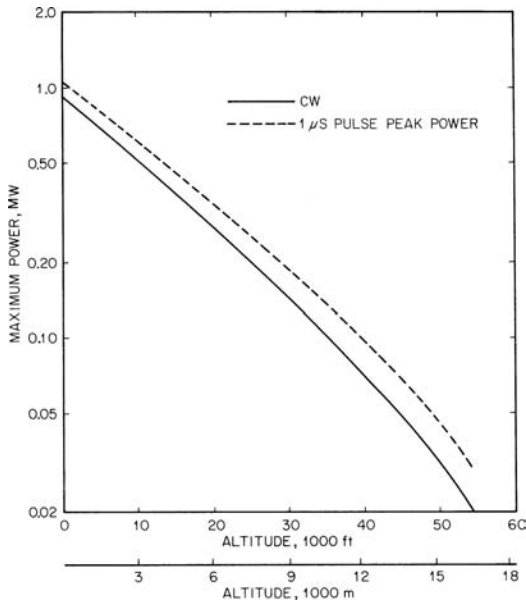


FIGURE 9-20 Power-handling capacity of WR-90 waveguide at 9.375 GHz

The electric field strength E across the slot aperture for a given slot-radiated power P is needed to compute the slot power-handling capacity. The electric field strength across the slot aperture is approximated by

$$E = (\eta_0 / b') \sqrt{P / 2R} \quad (9-44)$$

where R is the complementary dipole driving-point resistance. The baseline slot at 9.375 GHz is used to illustrate the slot power-handling capability. By combining Gilden and Gould's results³⁰ with Eq. 9-43, the maximum continuous-wave and peak power versus the altitude for the baseline slot is computed and shown in Figure 9-21. Equation 9-43 should be used for temperatures other than room temperature.

The power-handling capability can be improved by (1) avoiding capacitive windows and sharp corners in the waveguide manifold, (2) increasing the slot width, and (3) cooling and pressurizing the antenna system. The most effective technique for increasing power-handling capability is to pressurize the waveguide network or just certain waveguide sections that handle the largest power.³¹ If a slot-antenna array is wholly encapsulated in a pressurized housing, the increase in power-handling capability is proportional to the achievable pressure. Occasionally gases, such as nitrogen, Freon, and sulfur hexafluoride (SF_6), are used for pressurization, further increasing power-handling capability through the higher dielectric strength of the gases. The advantage of pressurized systems is that the pressure also keeps out moisture, dust, insects, and other debris that could cause arcing. This approach adds to the complexity of the mechanical design and fabrication of the antenna due to seals and other added equipment. Pressurization does add to the complexity and cost of any design.

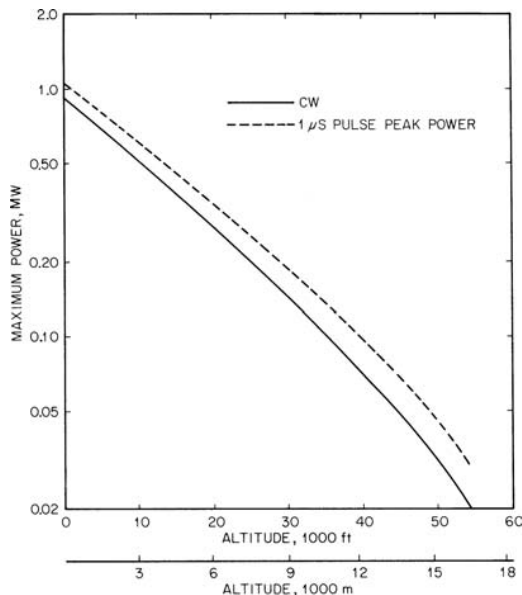


FIGURE 9-21 Power-handling capacity of the baseline slot in the planar array

9.7 TOLERANCE AND FABRICATION TECHNIQUES

In practice an antenna will include errors introduced by inaccuracies in the manufacturing process. Manufacturing costs are higher when excessively tight tolerances must be maintained. Thus a systematic allocation of dimensional tolerances is usually a worthwhile part of the design process.

The tolerance for waveguide width can be established by computing the incremental phase per unit change in the manifold and feed waveguide width. To determine the required dimensional tolerance for slots, the incremental conductance (or impedance) per unit change in the slot length from the resonant length is needed. These incremental values per unit change of slot parameters can be determined from given slot parameters discussed in preceding sections. Listed in Table 9-1 are the incremental phases of baseline shunt and series slots and the *S*-band edge slot due to the decrease of 1 percent slot length from resonant lengths. Note that the tolerance requirement for edge-slot arrays is less critical than for other slot arrays.

The difference in the incremental phase change indicates that the mutual-coupling effects reduce the longitudinal shunt-slot frequency-bandwidth performance and increase tolerance requirements. Sidelobe performance can be greatly degraded by a 1 percent variation in the slot length in a planar array of longitudinal shunt slots.

Fabrication of the commonly used edge-slot array is quite different from that of other slot arrays. Since the edge slot is cut into the narrow wall and wrapped around the broad faces of a rectangular waveguide, an edge-slot array can be easily fabricated by cutting (sawing) the slots at the proper angle and depth. The results are, in general, quite good. To generate a planar edge-slot array, slotted-waveguide sections are manifolded and assembled side by side to form a planar aperture.

To fabricate a planar array of longitudinal shunt slots or edge-wall folded shunt slots, a milling machine may be used to cut the slot on a thin, flat plate (face plate) such as shown in Figure 9-18. The back plate is milled out of a thick metal plate. Two narrow walls of the shunt and series waveguides are fabricated on the opposite sides of the metal plate, and the series slots are cut on the common walls of the shunt and series guides. Assembling the back plate to the face plate forms the shunt waveguides. Attaching the feed covers to the back plate forms the series waveguides.

The common techniques for assembling the face plate to the back plate array of broad-wall slots are

- **Welding** Both laser welding and electron-beam welding can be used to join the face plate to the back plate. Electron-beam welding has been successfully applied to manufacture antennas made of magnesium. Because of the inherent properties of aluminum, applying the electron-beam welding technique to this metal in assembling the flat-plate antenna requires additional investigative effort. Laser welding of the face plate to the back plate is possible, but the process is better suited to welding very thin metal cross sections and has limited application to flat-plate antennas.

TABLE 9-1 Incremental Phase Change of Slot with 1 Percent Change in Slot Length from Resonant Length

Longitudinal shunt slot		Edge slot	
Single	Coupled	Rotational series slot	Coupled
8.1°	19.3°	17.1°	2.7°

- **Brazing** Dip brazing has been widely used to secure two aluminum objects together. Many aluminum slotted-waveguide arrays are assembled by using dip brazing. The disadvantages of this technique are that the residue deposits left in the waveguides and on the slots, and the overall material shrinkage must be accounted for. The fluxless-brazing technique for joining the aluminum face and back plates together in a retort with inert atmosphere has been investigated with successful results.³²
- **Bonding** There are three commonly used bonding techniques: (a) thin-film bonding, (b) conductive-epoxy bonding, and (c) diffusion bonding. Thin-film bonding and conductive-epoxy bonding are quite popular in flat-plate-antenna applications.

Other techniques such as screw fasteners and twist tabs have been employed to join the face and back plates together. These two methods are either limited as to application or less accurate, adversely affecting antenna performance when compared with the preceding three assembling techniques. These approaches are prone to RF leakage.

The linear array is subject to fewer tolerance and assembly problems than the planar array. To fabricate and assemble a planar array such as the flat-plate antenna, both before and after the process of joining the face and back plates together, attention must be paid to keep the antenna aperture flat and the plates aligned accurately. A $0.01\lambda_0$ warping of an antenna can significantly degrade sidelobe performance, and a $0.007\lambda_0$ misalignment of the front and back plates will produce a grating-lobe level higher than -20 dB. Slot-length tolerance depends on the desired radiation performance.

Although a slot-antenna array can be designed and fabricated by using the design information given in this chapter, antenna performance may be less than expected. Some of the second-order effects, such as the slot alternating displacements, alternating inclined angles, edge effects, nonuniform aperture illumination, slot-length errors, and internal coupling through higher-order waveguide modes have not been addressed. The combined effects of neglecting second-order effects, manufacturing tolerances, and inaccurate manifold design usually significantly degrade practical slotted-array performance. For instance, a -35 -dB sidelobe-level array design may have only -28 -dB sidelobes in practice. Thus a thorough error analysis is recommended to define dimensional tolerances and to determine aperture illumination errors. Normally, overdesign is used to allow for design and manufacturing deficiencies, the degree of overdesign usually being determined by experience. Note that some of the degradation factors will be more noticeable for a -30 -dB sidelobe-level array design than for a -20 -dB design.

REFERENCES

1. T. Takeshima and Y. Isogai, "Frequency Bandwidth of Slotted Array Aerial System," *Electron. Engg.* (February 1969): 201–204.
2. D.-C. Chang, C.-G. Jan, and Y.-H. Yang, "A Semi-Theoretical Design Procedure for the Linear Edge Slot Arrays," *Microwave Conf. Proc. 1993, APMC '93, Asia-Pacific*, vol. 1 (1993): 1–78, 1–82.
3. V. V. S. Prakash, S. Christopher, and N. Balakrishnan, "Method-of-Moments Analysis of the Narrow-Wall Slot Array in a Rectangular Waveguide," *Microwaves, Antennas Propagat., IEE Proc.*, vol. 147, issue 3 (June 2000): 242–246.
4. L. A. Kurtz and J. S. Yee, "Second Order Beams of Two-Dimensional Slot Arrays," *IRE Trans. Antennas Propagat.*, vol. AP-5 (1957): 356–362.
5. R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, Part I, Chap. 14 (New York: McGraw-Hill, 1969).
6. H. M. Altschuler and A. A. Oliner, "Microwave Measurements with a Lossy Variable Short Circuit," Res. Rep. R-399-54, Polytechnic Institute of Brooklyn, Brooklyn, NY, 1954.

7. R. J. Stevenson, "Theory of Slots in Rectangular Waveguides," *J. App. Phys.*, vol. 19 (1948): 24–38.
8. A. A. Oliner, "The Impedance Properties of Narrow Radiating Slots in the Broad Face of Rectangular Waveguide," Parts I and II, *IRE Trans. Antennas Propagat.*, vol. AP-5 (1957): 4–20.
9. H. Y. Yee, "Impedance of a Narrow Longitudinal Shunt Slot in a Slotted Waveguide Array," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (1974): 589–592.
10. L. Lewin, *Advanced Theory of Waveguides* (London: Iliffe & Sons, Ltd., 1951): 106–114.
11. N. H. Frank and L. J. Chu, MIT Rad. Lab. Rep. 43-6 and 43-7, 1942.
12. J. P. Montgomery, unpublished work; private communication.
13. Vu Khac Thong, "Impedance Properties of a Longitudinal Slot Antenna in the Broad Face of a Rectangular Waveguide," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (1973): 708–710.
14. R. W. Lyon and A. J. Sangster, "Efficient Moment Method Analysis of Radiating Slots in a Thick Walled Rectangular Waveguide," *IEE Proc. (London)*, vol. 128, part H (1981): 197–205.
15. C.-G. Jan, P. Hsu, and R.-B. Wu, "Moment Method Analysis of Sidewall Inclined Slots in Rectangular Waveguides," *IEEE Trans. Antennas Propagat.*, vol. AP-39 (January 1991): 68–73.
16. H. Y. Yee and P. Stelitano, "Narrow Wall I-Slot Characteristics," *IEEE Trans. Antennas Propagat.*, vol. AP-40 (February 1992): 224–228.
17. K. Falk, "Admittance of a Longitudinal Slot in Ridge Waveguide," *IEE Proc. (London)*, vol. 135, part H (1988): 263–268.
18. W. H. Watson, "Resonant Slots," *J. IEE (London)*, part IIIA, vol. 93 (1946): 747–777.
19. R. J. Stegen, "Longitudinal Shunt Slot Characteristics," Hughes Aircraft Co. Tech. Mem. 261, Culver City, CA, November 1951.
20. I. Kaminow and R. J. Stegen, "Waveguide Slot Array Design," Hughes Aircraft Co. Tech. Mem. 348, Culver City, CA, 1954.
21. S. W. Lee and W. R. Jones, "On the Suppression of Radiation Nulls and Broadband Impedance Matching of Rectangular Waveguide Phased Array," *IEEE Trans. Antennas Propagat.*, vol. AP-19 (1971): 41–51.
22. G. J. Stern and R. S. Elliott, "Resonant Length of Longitudinal Slots and Validity of Circuit Representation: Theory and Experiment," *IEEE Trans. Antennas Propagat.*, vol. AP-33 (November 1985): 1264–1271.
23. R. S. Elliott, "An Improved Procedure for Small Arrays of Shunt Slots," *IEEE Trans. Antennas Propagat.*, vol. AP-31 (January 1983): 48–53.
24. R. J. Chignell and J. Roberts, "Compact Resonant Slot for Waveguide Arrays," *IEE Proc. (London)*, vol. 125, no. 11 (November 1978): 1213–1216.
25. T. Shicopoulos, "C-Slot: A Practical Solution for Phased Arrays of Radiating Slots Located on the Narrow Side of Rectangular Waveguides," *IEE Proc. (London)*, vol. 129, part H (April 1982): 49–55.
26. R. S. Elliott, *Antenna Theory and Design*, Chap. 8 (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1981).
27. R. S. Elliott and L. A. Kurtz, "The Design of Small Slot Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-26 (1978): 214–219.
28. P. S. Carter, "Circuit Relations in Radiating System and Applications to Antenna Problems," *IRE Proc.*, vol. 20 (1932): 1004–1041.
29. M. G. Chernin, "Slot Admittance Data at Ka Band," *IRE Trans. Antennas Propagat.*, vol. AP-4 (1956): 632–636.
30. M. Gilden and L. Gould, *Handbook on High Power Capabilities of Waveguide Systems* (Burlington, MA: Microwave Associates, Inc., June 1963).
31. J. Ciavolella, "Take the Hassle out of High Power Design," *Microwaves* (June 1972): 60–62.
32. "Aluminum Brazed Antenna Manufacturing Methods," AFML Cont. F33615-75-5266, Antenna Dept., Hughes Aircraft Company, Culver City, CA, 1975.

Chapter 10

Surface-Wave
Antennas*

Francis J. Zucker

*Rome Laboratory
Hanscom Air Force Base*

CONTENTS

10.1	INTRODUCTION	10-2
10.2	PROPERTIES AND MEASUREMENT OF SURFACE WAVES	10-2
10.3	SURFACE-WAVE ANTENNAS: DESIGN PRINCIPLES	10-8
10.4	SURFACE-WAVE ANTENNAS: SPECIFIC STRUCTURES	10-19
10.5	LONG-WIRE ANTENNAS	10-23

* This chapter was updated by William F. Croswell, Harris Corporation. The original chapters were written by Francis J. Zucker (as modified by Richard C. Johnson) and by Edmund A. Laport.

10.1 INTRODUCTION

The two types of traveling-wave antennas discussed in this chapter are illustrated in Figure 10-1. In Figure 10-1a, a *surface wave* (also called a *trapped wave* because it carries its energy within a small distance from the interface) is launched by the feed F and travels along the dielectric rod to the termination T . Since a surface wave radiates only at discontinuities, the total pattern of this antenna (normally end-fire) is formed by the interference between the feed and terminal patterns.¹ The dielectric material could alternatively be an artificial one, e.g., a series of metal disks or rods (the Yagi-Uda antenna). On the other hand, discontinuities can be placed all along the surface-wave structure, producing radiation in the manner of a slotted waveguide, except that the wave propagates on an open, not a shielded, guide. The array of dipoles proximity-coupled to a two-wire transmission line in Figure 10-1b illustrates this second antenna type; here we are evidently using the term *surface wave* loosely to designate any wave mode that propagates along an open interface without radiating.

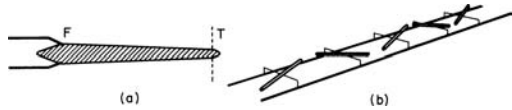


FIGURE 10-1 Surface-wave-antenna types: (a) dielectric rod and (b) array of proximity-coupled dipoles (thin lines are dielectric supports)

Two closely related antenna structures are discussed in other chapters: the helix (which involves a surface wave in one of its modes of operation—Chapter 12) and the frequency-independent antennas (Chapter 13). The short Yagi-Uda, in whose design surface waves play no role, is covered in Chapter 3. An additional discussion of surface wave antennas can be found in Zucker.²

Surface-wave antennas lend themselves to flush, or at least low-silhouette, installation. Their gain normally does not exceed 20 dB, and control over their pattern shape (including sidelobes) is *limited*. Bandwidth is generally narrow. Frequency coverage extends from the popular Yagi-Udas at very high frequencies (VHF) to printed microwave (and millimeter-wave) antennas. By contrast, arrays excited by surface waves produce patterns that can in principle produce as high gain and be controlled as accurately as the patterns of nonresonant slot arrays.

10.2 PROPERTIES AND MEASUREMENT OF SURFACE WAVES

For design purposes, a surface wave is sufficiently characterized by one of its parameters, for example, the wavelength in the direction of propagation along the interface. The various parameters will now be defined and formulas given for their interrelationship; it will then be shown how they can be calculated if the surface impedance is known or how they can be determined by measurements.

Interrelationship of Parameters

The governing relation is the separability condition for the wave equation. In rectangular coordinates (see Figure 10-2a),

$$k_x^2 + k_y^2 + k_z^2 = k^2 \quad (10-1)$$

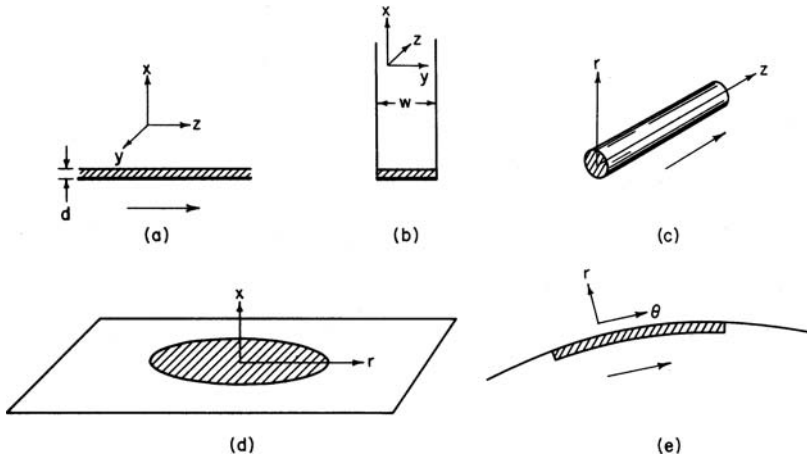


FIGURE 10-2 Geometries for surface-wave propagation over dielectric sheets: (a) Infinite-plane sheet-on-metal, (b) Rectangular duct, (c) Axial-cylindrical rod, (d) Radial-cylindrical sheet-on-metal, (e) Azimuthal-cylindrical or spherical cap. Heavy lines indicate metal surfaces. Arrows point in the direction of propagation.

where the k 's, called *wave numbers*, are in general complex—for example, $k_x = \beta_x - j\alpha_x$, with β_x the phase constant in radians per unit length (inches, centimeters, or meters), and α_x the attenuation in nepers per unit length. The wave number k is that of the medium in which the wave travels, $k = \omega\sqrt{\epsilon_0\mu_0}$, with ω the angular frequency, ϵ_0 the dielectric constant, and μ_0 the permeability. In air, k is pure real; it is related to the wavelength λ by $k = \beta = 2\pi/\lambda$ and to the phase velocity of light c by $k = \omega/c$. Similarly, the phase constant β_z along the surface is related to the surface wavelength λ_z by $\beta_z = 2\pi/\lambda_z$ and to the surface phase velocity v_z by $\beta = \omega/v_z$. The following ratios are therefore equivalent:

$$\frac{\beta_z}{k} = \frac{\lambda}{\lambda_z} = \frac{c}{v_z} \quad (10-2)$$

When this ratio is greater than 1, we speak of a “slow” wave (i.e., slower than light, because $v_z < c$); when less than 1, of a “fast” wave ($v_z > c$). The surface wavelength λ_z of slow waves is shorter than λ ; that of fast waves, longer than λ (as in shielded waveguide).

A *surface* wave is one that propagates parallel to the interface and decays vertically to it; that is, the phase constant β_x is zero, and $k_x = -j\alpha_x$ (pure attenuation), with α_x positive. Assume first that the wave extends indefinitely in the transverse direction so that $k_y = 0$ (see Figure 10-2a). Equation 10-1 requires that k_z be pure real; that is, there can be no attenuation in the direction of propagation, and we obtain the simple but basic relation

$$\beta_z^2 = k^2 + \alpha_x^2 \quad (10-3a)$$

or equivalently, by using Eq. 10-2,

$$\frac{\lambda}{\lambda_z} = \frac{c}{v_z} \sqrt{1 + \left(\frac{\alpha_x \lambda}{2\pi}\right)^2} \quad (10-3b)$$

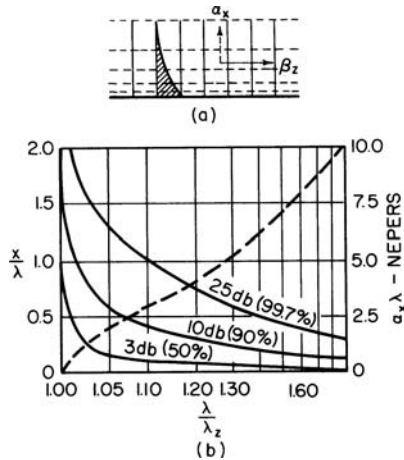


FIGURE 10-3 Phase and amplitude contours of a plane surface wave. (a) Solid lines indicate constant-phase fronts; dashed lines, constant-amplitude fronts; shaded region shows amplitude decay. (b) Solid lines indicate height x/λ above surface of 3-dB, 10-dB, and 25-dB constant-amplitude or power contours as a function of relative surface wavelength λ/λ_z (use left-hand ordinate); numbers in parentheses refer to the percentage of total energy carried by the surface wave between the interface and each contour. The dashed line indicates the relation between relative surface wavelength and vertical attenuation $\alpha_x \lambda$ according to Eq. 10-3b (use right-hand ordinate).

from which it follows that this is a slow wave. Equation 10-3b is plotted as the dashed line in Figure 10-3b. Slower-than-light surface waves are used chiefly in end-fire antenna design (see Sections 10.3 and 10.4). Figure 10-3a illustrates constant-phase and constant-amplitude fronts, which are at right angles to each other provided only that the medium above the interface is lossless. (Ohmic losses in the surface produce a slight forward tilt in the phase front and a small α_x .) The more closely the surface phase velocity approaches that of light, the smaller is α_x (see Eq. 10-3b) and the larger, therefore, the vertical extent of the surface wave in Figure 10-3a.

All components of the total electromagnetic field of the surface wave, for example, E_z , are of the form

$$E_z(x, z, t) = E_z e^{-\alpha_x x} e^{-j\beta_z z} e^{j\omega t} \quad (10-4)$$

The vertical decay of the E_z , E_x , and H_y components of a TM surface wave ($H_z = 0$) is shown in Figure 10-4a, and the composite electric field lines over a full wavelength interval (longitudinal section) in Figure 10-4b. It can be shown that E_z is in phase with H_y , E_x in phase quadrature. The first two components therefore carry all the power along the interface, while E_x and H_y form a vertically pulsating storage field. The dielectric slab-on-metal below the interface in Figure 10-4 is used as an example. Since the form of Eq. 10-4, and thus of the surface-wave field lines, does not depend on the detailed structure of the medium below the interface, a multilayer dielectric or a corrugated sheet, etc., could have been shown equally well.

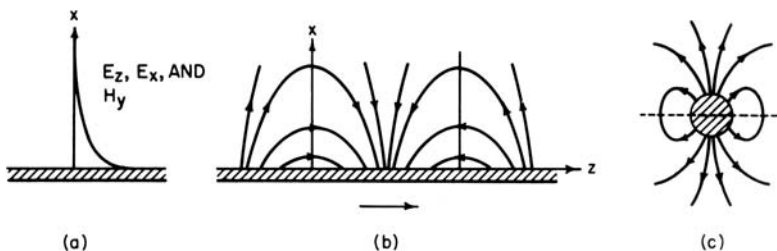


FIGURE 10-4 Surface-wave field structure above interface. (a) TM-wave components on plane surface. (b) Composite electric field lines over a one-wavelength interval (arrow points in the direction of propagation). (c) Composite electric field lines of HE_{11} wave on an axial-cylindrical surface (cross section). Dashed line shows the electric-image plane (see Section 10.4).

The field (or power) decay in decibels is plotted as a function of height x/λ above the surface in Figure 10-3b (based on Eq. 10-3a). If we assume a relative surface wavelength λ/λ_z (or relative phase velocity c/v_z) equal to 1.1, for example, the field at height $x = \lambda$ has decayed by 25 dB below its value at the interface.

Assume next that two parallel walls are erected normal to the surface, forming a duct in the z direction (see Figure 10-2b). (Examples other than the dielectric-loaded channel in this figure are the channel with a corrugated bottom and the trough guide.) The transverse wave number is fixed by the duct width w as in conventional waveguide theory; from Eq. 10-1 we now have, instead of Eq. 10-3a,

$$\beta_z^2 = k^2 + \alpha_x^2 - \left(\frac{n\pi}{w} \right)^2 \quad n = 0, 1, 2, \dots \quad (10-5a)$$

If $n = 0$ (no variation in y direction), Eqs. 10-3, and therefore also Figure 10-3b, still apply. If $n = 1$ (half a sinusoid in y direction),

$$\frac{\lambda}{\lambda_z} = \frac{c}{v_z} = \sqrt{1 + \left(\frac{\alpha_x \lambda}{2\pi} \right)^2 - \left(\frac{\lambda}{2w} \right)^2} \quad (10-5b)$$

Since the second term is usually smaller than the third, it follows that the wave is usually, though not necessarily, faster than light. Fast as well as slow surface waves can be used to excite arrays of discrete elements. The field and power decay still follows Figure 10-3a and b, provided that the abscissa in Figure 10-3b is relabeled $\sqrt{(\lambda/\lambda_z)^2 + (\lambda/2w)^2}$ to account for the difference between Eqs. 10-3b and 10-5b. The heights x/λ corresponding to the 25-dB contour in Figure 10-3b are useful for estimating the extent to which the duct walls, which theoretically must be infinite in height, may be lowered without perturbing the surface wave.

Surface-wave geometries other than rectangular are shown in Figure 10-2c-e. The separability condition for the wave equation in axial- and radial-cylindrical coordinates (Figure 10-2c and d) reads

$$k_r^2 + k_z^2 = k^2 \quad (10-6)$$

(note that k_θ does not appear), which is formally identical with Eq. 10-1 for the case $k_r = 0$. The interrelationships of the surface-wave parameters given in Eqs. 10-3 still hold if k_r is substituted for k_z . Above the surface, the field components decay like Hankel functions; in the axial-cylindrical case, for example,

$$E_z(r, z, t) = E_z H_0^{(2)}(\alpha_r r) e^{-j\beta_z z} e^{j\omega t} \quad (10-7)$$

The circumferential field dependence in the case of the HE_{11} mode—the one of principal interest in this geometry—is one full sinusoid. Composite E lines for this *hybrid* wave (so called because of the presence of both E_z and H_z) are shown in Figure 10-4c. [The structure inside the circular interface could equally well be a metal rod with a dielectric mantle, an array of circular disks (cigar antenna), or even an array of dipoles (Yagi-Uda).] For large radii, the Hankel function is asymptotic to the exponential, and Eq. 10-7 differs negligibly from Eq. 10-4. As a consequence, the 25-dB contours in Figure 10-3b still hold and the 10-dB contours nearly so; these curves are useful in estimating the degree of coupling between adjacent end-fire line sources. In the radial-cylindrical case, the field components decay exponentially away from the surface so that Figure 10-3b applies exactly. But the constant-amplitude fronts will now no longer be parallel to the surface because the field components decay radially like Hankel functions (asymptotically like $1/\sqrt{r}$), bringing the fronts closer to the surface with increasing r .

In the azimuthal-cylindrical or spherical case, it can be shown that the wave numbers are always complex so that these surface waves are no longer completely trapped. Because the radius of curvature is usually large, it is convenient to treat azimuthal waves as a perturbation of Eqs. 10-3 for flat surfaces.

Calculation of Surface-Wave Parameters

Surface waves are modes on unshielded waveguides. As in the case of shielded waveguides, the boundary-value problem is often conveniently stated in terms of the *transverse resonance condition*, which requires that the equivalent network of a transverse section through the waveguide be resonant.³ With Z_x the impedance looking straight up from the interface and Z_s the surface impedance looking straight down, the resonance condition reads

$$Z_x + Z_s = 0 \quad (10-8)$$

Z_x depends on the geometry and the mode; in rectangular coordinates,

$$Z_x = \frac{k_x}{\omega \epsilon_o} \quad \text{for TM waves} \quad (10-9a)$$

$$Z_x = \frac{\omega \mu_o}{k_x} \quad \text{for TE waves} \quad (10-9b)$$

In the case of a surface wave, $k_x = -j\alpha_x$; Z_x is therefore capacitive in Eq. 10-9a and inductive in Eq. 10-9b. It follows from Eq. 10-8 that the TM surface wave requires an inductive surface, $Z_s = jX_s$, and the TE surface wave a capacitive surface, $Z_s = -jX_s$ (X_s positive). Equation 10-8 now reads

$$\frac{\alpha_x}{k} = \frac{X_s}{R_c} \quad \text{for TM waves} \quad (10-10a)$$

$$\frac{\alpha_x}{k} = \frac{R_c}{X_s} \quad \text{for TE waves} \quad (10-10b)$$

where $R_c = \sqrt{\mu_o \epsilon_o} = 377 \Omega$, β_z or λ / λ_z follows immediately from Eqs. 10-3.

Knowledge of the surface impedance thus solves the boundary-value problem. It is instructive to calculate this impedance for two simple cases: surface waves on a dielectric slab-on metal and on a corrugated-metal sheet. Looking down from the air-dielectric interface in Figure 10-2a, we see a dielectric-loaded region short-circuited at the end. By using transmission-line theory, assuming that the wave is TM, and noting that a wave that is TM in the direction of propagation is also TM transverse to it,

$$X_s = \frac{k_x}{\omega \epsilon_1} \tan k_x d \quad (10-11)$$

where k_x is the vertical wave number in the slab of dielectric constant ϵ_1 . Since X_s must be positive real, $k_x = \beta_x$. (We could now write the field components below the interface in the form of Eq. 10-4, with a sinusoidal variation in x replacing the exponential decay; the field configuration below the interface, however, is rarely of interest to the engineer.) Substitution of Eq. 10-11 in Eq. 10-10a gives one equation in two unknowns, α_x and β_x . α_x is a function of β_z through Eq. 10-3a; β_x can also be expressed as a function of β_z , since Eq. 10-1 reads

$$\beta_x^2 + \beta_z^2 = k_1^2$$

in the dielectric medium ($k_1 = \omega \sqrt{\mu_1 \epsilon_1}$). The solution⁴ shows that the lowest TM mode has no cutoff and that the phase velocity of the surface wave lies between c and $c\sqrt{\epsilon_1/\epsilon_0}$, approaching the free-space velocity in air for very thin sheets and the free-space velocity in the dielectric for thick sheets (numerical results are in Section 10.4). The analysis for a TE wave uses Eq. 10-10b and results in a lowest mode that does have a cutoff. It is worth pointing out that below cutoff a surface wave does not become evanescent but ceases to exist altogether.

The interface of the corrugated-metal sheet (see Figure 10-5) is an imaginary plane through the top of the teeth. X_s is zero over the teeth and equals $R_c \tan kd$ over the groove, which are short-circuited parallel plate transmission lines (inductive in the range $0 < kd < \pi/2$). If there are many grooves per wavelength, the surface impedance over the teeth and grooves may be averaged; that is, $X_s \equiv [g/(g+t)]R_c \tan kd$. Substitution in Eq. 10-10a gives one equation in one unknown. The results of this approximation turn out to be good provided there are at least five grooves per wavelength. A more exact calculation takes into account the higher-order (below-cutoff) modes at the mouth of the grooves, which produce higher-order waves along the interface that modify the field lines sketched in Figure 10-4a and *b* only in the close vicinity of the teeth.

Measurement of Surface-Wave Parameters

In the case of surface waves, a vertical metal plate large enough to reflect at least 90 percent of the incident energy (use Figure 10-3b to determine the minimum size) produces deep nulls



FIGURE 10-5 Corrugated-metal surface. Arrow points in the direction of propagation.

at half-wavelength intervals along the surface. Because of turbulence in the aperture field near the feed, measurements should be made at a minimum distance from the feed given by

$$d_{\min} \cong \frac{0.17\lambda}{(\lambda/\lambda_z) - 1}$$

If the frequency is so low that reflector size becomes a problem, the distance between nulls is found by feeding the probe signal into a phase-comparison circuit (using magic T and attenuator) that measures the traveling-wave phase with respect to a reference signal.

The probe used in these measurements should have a small cross section so as not to perturb the field. It needs chokes to suppress antenna currents on its outer conductor, and it must couple only into the E or the H field, with good discrimination against the unwanted one. A coaxially fed monopole or a waveguide horn pinched at the mouth (with dielectric loading to keep it above cutoff)⁵ is very satisfactory; loops and horizontal dipoles, less so. The distance between probe and aperture should be as small as possible without perturbing the aperture field, and it should be held constant by providing a rigid probe carriage.

10.3 SURFACE-WAVE ANTENNAS: DESIGN PRINCIPLES

Radiation of Surface-Wave Antennas

Before presenting design principles (for optimum gain, beamwidth, sidelobe level, and bandwidth), it is useful to examine how surface-wave antennas radiate. Two typical structures are shown in Figure 10-6. The feed F (consisting of a monopole and reflector in Figure 10-6a and of a circular or rectangular waveguide in Figure 10-6b) couples a portion of the input power into a surface wave, which travels along the antenna structure to the termination T , where it radiates into space. The ratio of power in the surface wave to total input power (efficiency of excitation) is usually between 65 and 75 percent. Power not coupled into the surface wave is directly radiated by the feed in a pattern resembling that radiated by the feed when no antenna structure is in front of it.

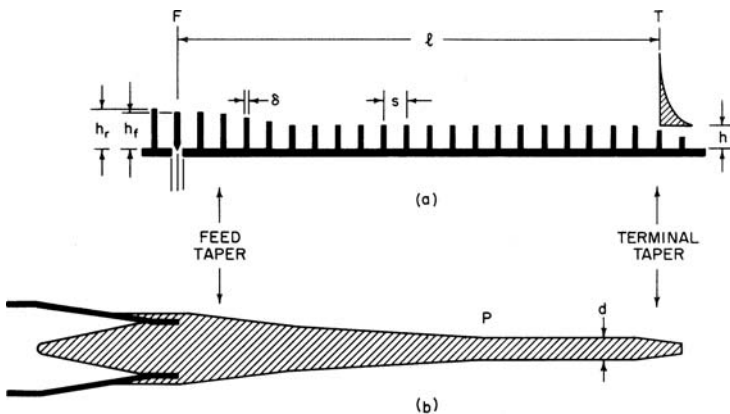


FIGURE 10-6 Surface-wave antenna structures: (a) Yagi-Uda (row of monopoles) on ground plane, excited by dipole-reflector combination; (b) Dielectric rod, excited by circular or rectangular waveguide (broad wall in plane of paper)

The tapered regions in Figure 10-6 serve diverse purposes. The *feed taper* increases the efficiency of excitation and also affects the shape of the feed pattern. The *body taper* (extending to point *P* in Figure 10-6*b*) suppresses sidelobes and increases bandwidth. Because a reflected surface wave spoils the pattern and bandwidth of the antenna, a *terminal taper* is employed to reduce the reflected surface wave to a negligible value.

The surface wave illuminates the terminal aperture (plane perpendicular to the antenna axis through *T*) in a circular region whose radius increases as the transverse attenuation of the surface wave diminishes or, equivalently, as the surface phase velocity approaches that of light (Eq. 10-3*b*). The larger the illuminated region, the higher the gain produced in the radiation pattern. Since the surface-wave phase front lies in the aperture plane, the pattern peaks in the end-fire direction. A simple but only approximate expression^{2,6} for the terminal radiation pattern $T(\theta)$ is

$$T(\theta) \cong \frac{1}{(\lambda/\lambda_z) - \cos \theta} \quad (10-12)$$

where θ is the angle off end-fire and λ_z is the surface wavelength at *T* (just to the left of the terminal taper, to be precise). This pattern has neither nulls nor sidelobes and, as expected, falls off more rapidly the closer λ/λ_z is to 1. More accurate expressions^{2,7-9} and an experimental determination¹⁰ indicate that the pattern is actually about 20 percent narrower than in Eq. 10-12. The dashed line in Figure 10-7 shows this corrected terminal radiation pattern for $\lambda/\lambda_z = 1.08$, which is the optimum relative phase velocity for a maximum-gain antenna 4λ long (Eq. 10-14*b*).

The radiation pattern of the feed shown in Figure 10-6*a* is quite broad (approximately a cardioid in polar coordinates). In combination with the terminal radiation $T(\theta)$, it produces the total surface-wave antenna pattern, shown by the solid line in Figure 10-7 (experimental data for antenna length $\ell = 4\lambda$, $\lambda/\lambda_z = 1.08$). In the vicinity of end-fire, the terminal pattern predominates; as θ increases, interference with the feed pattern first narrows the beam and then produces a sidelobe at 35° ; beyond 45° the feed pattern predominates. Patterns of all surface-wave antennas adjusted for maximum gain look like those in Figure 10-7, except that for antennas longer than 4λ the pattern spreads less in θ (first sidelobe closer to end-fire than 35°) and for shorter structures it spreads more. The field beyond the first sidelobe falls off more rapidly the larger the end-fire gain of the feed radiation.

The curves in Figure 10-7 are for the *H* plane, in which the element pattern of the Yagi-Uda monopoles (see Figure 10-6*a*) is omnidirectional. In the *E* plane, $T(\phi)$ (ϕ is the angle off end-fire) is narrowed by multiplication with the *E*-plane element factor (a dipole pattern modified by the effects of mutual impedance), which for $\phi < 60^\circ$ approximates $\cos \phi$ and then decays more gradually (no null at $\phi = 90^\circ$). The total pattern in the *E* plane is therefore also narrower than in the *H* plane, and sidelobes are lower by 2 to 3 dB.

In the case of the HE_{11} mode on a dielectric rod (see Figure 10-6*b*), the *E*- and *H*-plane element factors, and consequently the beamwidths and sidelobe levels, are more nearly the same. The sidelobe level is now usually *higher* by 0.5 to 1.5 dB in the *E* plane than in the *H* plane, the precise amount depending on the relative shape of the *E*- and *H*-plane feed pattern.

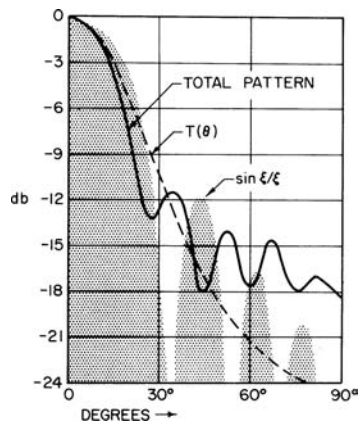


FIGURE 10-7 Radiation pattern of 4λ -long surface-wave antenna adjusted for maximum gain in accordance with Eq. 10-14*b* ($\lambda/\lambda_z = 1.08$). $T(\theta)$ is the *H*-plane pattern of the surface-wave distribution in the terminal plane. For comparison, $\sin \xi/\xi$ is also shown (shaded region).

Instead of being regarded as a combination of feed and terminal radiation, the surface-wave antenna pattern can be viewed as an integral over the field distribution along the antenna structure. What this distribution looks like can be seen in Figure 10-8: the amplitude $A(z)$ along the row of monopoles (see Figure 10-6a) starts with a hump near the feed and flattens out as the surface wave peels itself out of the near field of feed and feed taper. The surface wave is well established at a distance ℓ_{\min} from the feed where the radiated wave from the feed, propagating at the velocity of light, leads the surface wave by about 120° :

$$\ell_{\min} k_z - \ell_{\min} k = \frac{\pi}{3} \quad (10-13)$$

The location of ℓ_{\min} on an antenna designed for maximum gain is seen in Figure 10-8 to be about halfway between feed and termination. Since the surface wave is fully developed from this point on, the remainder of the antenna length is used solely to bring feed and terminal radiation into the proper phase relation for maximum gain. (In the absence of a feed taper, ℓ_{\min} occurs closer to F and the hump is higher.)

Were there no hump at all, the surface-wave antenna would radiate the familiar $\sin \xi/\xi$ pattern [$\xi = (\pi/\lambda)(\lambda/\lambda_z - \cos \theta)$] produced by constant-amplitude illumination. This pattern is shown for comparison in Figure 10-7 (shaded region); it is an unnatural pattern for parasitically excited antennas, which necessarily have feed turbulence.

Low-sidelobe and broadband designs require tapers that extend over a substantial part of the antenna (see Figure 10-6b). Although the field along one particular taper has been examined in detail,^{11,12} the effect of tapers on the total radiation pattern is not well understood. When the taper is very gradual, the phase velocity at each point has the value we would expect from the local surface impedance; when it is sharp, a leaky wave which fills in the minima of the pattern, widens the beam and reduces sidelobes is produced.

Design for Maximum Gain

The phase velocity along the antenna and the dimensions of the feed and terminal tapers in the maximum-gain design of Figure 10-6a must now be specified.

If the amplitude distribution in Figure 10-8 were flat, maximum gain would be obtained by meeting the Hansen-Woodyard condition (strictly valid only for antenna lengths $\ell \gg \lambda$), which requires the phase difference at T between the surface wave and the free-space wave from the feed to be approximately 180° :

$$\ell k_z - \ell k = \pi$$

or, equivalently,

$$\frac{\lambda}{\lambda_z} = 1 + \frac{\lambda}{2\ell} \quad (10-14a)$$

which is plotted as the upper dashed line in Figure 10-9.

Since the size and extent of the hump are a function of feed and feed-taper construction, the optimum terminal phase difference for a prescribed antenna length cannot easily be calculated.

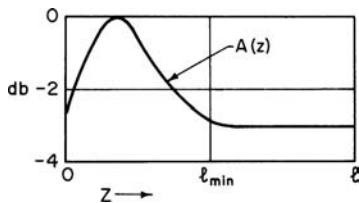


FIGURE 10-8 Amplitude $A(z)$ of field along the surface-wave antenna structure

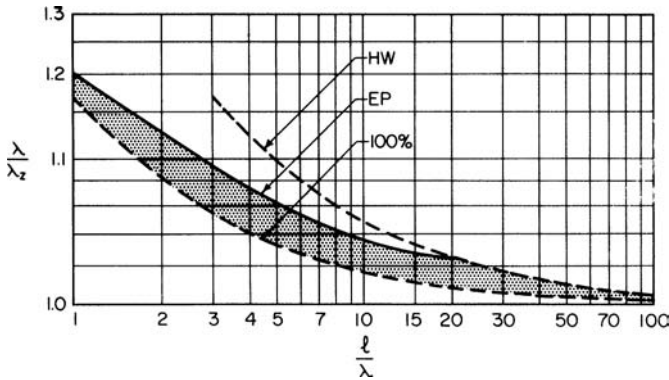


FIGURE 10-9 Relative-phase velocity $c/v = \lambda/\lambda_z$ for maximum-gain surface-wave antennas as a function of relative antenna length ℓ/λ . HW = Hansen-Woodyard condition, Eq. 10-14b; EP = Ehrenspeck and Poehler experimental values, Eq. 10-14b; 100 percent = idealized perfect excitation, Eq. 10-14c.

Experimental work on Yagi-Uda antennas (but without feed taper) by Ehrenspeck and Poehler¹³ and Ehrenspeck¹⁴ has shown that the optimum terminal phase difference lies near 60° for very short antennas, rises to about 120° for ℓ between 4λ and 8λ , then gradually approaches 180° for the longest antenna measured (20λ). Therefore

$$\frac{\lambda}{\lambda_z} = 1 + \frac{\lambda}{p\ell} \quad (10-14b)$$

with p starting near 6 for $\ell = \lambda$ and diminishing to approximately 3 for ℓ between 3λ and 8λ and to 2 at 20λ . This relation is plotted as the solid curve in Figure 10-9. In the presence of a feed taper, the optimum λ/λ_z values lie slightly below this curve.

If the efficiency of excitation is very high, interference between feed and terminal radiation is of minor importance, and the antenna need just be long enough so that the surface wave is fully established; that is, $\ell = \ell_{\min}$ in Figure 10-8. From Eq. 10-13,

$$\frac{\lambda}{\lambda_z} = 1 + \frac{\lambda}{6\ell} \quad (10-14c)$$

which is plotted as the lower bound of the shaded region in Figure 10-9.

Because feeds are more efficient when exciting slow surface waves than when the phase velocity is close to that of light, the solid line starts near the lower bound and ends at the upper. As long as λ/λ_z falls within the shaded region, its precise adjustment is not critical; with $\ell = 4\lambda$ and in the presence of a feed taper, the gain drops only 1 dB if λ/λ_z lies on the dashed lower bound instead of just below the solid upper bound.

Although this technique for maximizing the gain has been strictly verified only for Yagi-Uda antennas, data available in the literature on other structures suggest that optimum λ/λ_z values lie on or just below the solid curve in all instances. To produce these optimum values, the dependence of λ/λ_z on the structural parameters of the antenna must be calculated, measured, or sought in the literature.

The feed taper should begin at F with λ/λ_z between 1.2 and 1.3 and extend over approximately 20 percent of the full antenna length. Its exact shape is not important as long as a smooth transition is made to the main body of the antenna. The terminal taper should be approximately 0.5λ long, and to match the surface wave to space, λ_z at the end

of the taper should be as close as possible to λ . In the case of a polyrod ($\epsilon = 2.56$), for example, Figure 10-16 shows that λ_z is reasonably close to λ when

$$d/\lambda = 0.23$$

and the end of the rod is therefore blunt rather than pointed.

The *gain* above an isotropic radiator of a long ($\ell \gg \lambda$) uniformly illuminated end-fire antenna whose phase velocity satisfies Eq. 10-14a was shown by Hansen and Woodyard to be approximately

$$G \equiv \frac{7\ell}{\lambda}$$

If the design is based on the optimal phase velocities and taper dimensions just described, the gain for ℓ between 3λ and 8λ is

$$G \equiv \frac{10\ell}{\lambda} \quad (10-15)$$

which is 1.5 dB above the Hansen-Woodyard gain. For shorter lengths the gain may be 30 percent higher; for longer lengths, the proportionality factor slowly decreases because of ohmic loss and the difficulty of designing an efficient feed. Maximum gains reported in the literature are plotted in decibels as a function of ℓ/λ in Figure 10-10 (solid-line margin of shaded region marked "Gain"). These gains involve a slight degree of superdirectivity but not enough to produce excessively narrow bandwidths or high ohmic loss.

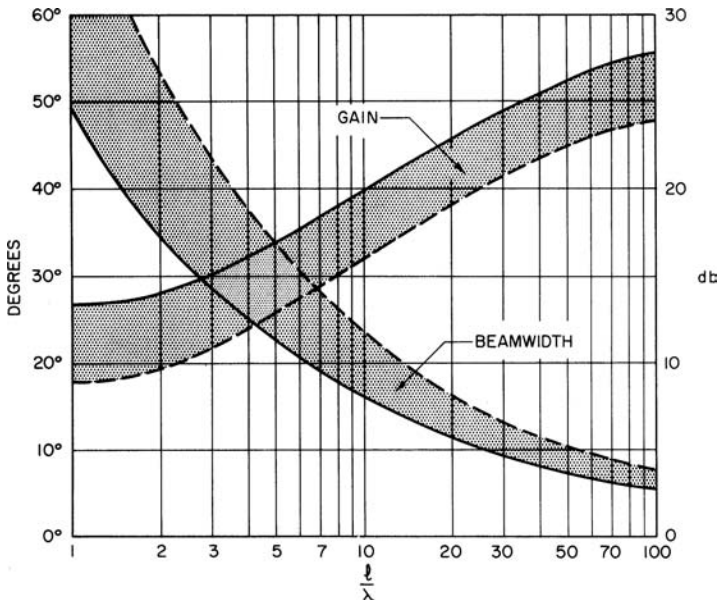


FIGURE 10-10 Gain and beamwidth of a surface-wave antenna as a function of relative antenna length ℓ/λ . For gain (in decibels above an isotropic source), use the right-hand ordinate; for beamwidth, the left-hand ordinate. Solid lines are optimum values; dashed lines are for low-sidelobe and broadband design.

The half-power *beamwidth* BW of a maximum-gain design is approximately

$$\text{BW} = 55 \sqrt{\frac{\lambda}{\ell}} \text{ degrees} \quad (10-16)$$

which lies just above the lower (solid-line) margin of the shaded region marked “Beamwidth” in Figure 10-10. Equation 10-16 gives an average figure, the beamwidth usually being slightly narrower in the *E* plane and slightly wider in the *H* plane.

The *sidelobe* level, for ℓ between 3λ and 8λ , is about 11 dB in the *H* plane and 10 or 14 dB in the *E* plane of a dielectric rod or Yagi-Uda, respectively (cf. above). For shorter antennas the sidelobes are somewhat higher; for longer antennas, lower.

The *bandwidth* within which the gain drops at most 3 dB is between ± 10 percent and ± 15 percent. Below the design frequency, gain slowly decreases as beamwidth widens; above it, the pattern deteriorates rapidly as the main beam splits and sidelobes rise.

By comparing Eq. 10-15 with the gain of a paraboloidal dish (equal sidelobe level is assumed), the surface-wave-antenna length is found to be related to the diameter d of an equal-gain dish by

$$\frac{\ell}{\lambda} \equiv \left(\frac{d}{\lambda} \right)^2 \quad (10-17)$$

This trade-off between a broadside and an end-fire structure becomes very disadvantageous to the latter as ℓ/λ increases, and surface-wave antennas longer than 10λ are rarely used. This limits the gain of surface-wave line sources in practice to 20 dB. (Arrays of line sources and end-fire area sources are practical for much higher gains.)

The gain predicted by Eq. 10-15 is for feeds of the type shown in Figure 10-6, with radiation patterns that are not very directive. Increasing the feed directivity increases the total antenna gain. For feeds whose radial extent is not much larger than the antenna cross section, it has been found experimentally¹⁵ that a 1- or 2-dB increase in the end-fire gain of the feed produces an approximately equal increase in the total antenna gain. If the feed is much larger than the antenna cross section, the gain continues to rise, though at a slower rate. A large horn aperture, corner reflector, or dish can add as much as 3 or 4 dB. Structures of this sort are part surface-wave, part aperture antenna.

The *backfire* antenna, which also combines these two features, produces a gain about 4 dB above that of Eq. 10-15.¹⁶ It consists of a surface-wave line source (for example, a Yagi-Uda, Figure 10-11) terminated by a flat circular plate P , which reflects the surface wave launched by the feed back toward F , where it radiates into space. Its radiation mechanism is not completely understood. The optimum phase velocity is chosen with the aid of the solid curve in Figure 10-9, but ℓ in this context must be reinterpreted as twice the physical antenna length ℓ' of the backfire antenna, since the surface wave traverses the antenna twice. By using Eq. 10-14b, the relation between surface phase velocity and ℓ' (for ℓ' between 1.5λ and 4λ) is $\lambda/\lambda_z = 1 + \lambda/6\ell'$. The feed taper should be much reduced in size or omitted. The plate diameter d is related to ℓ' by

$$\frac{d}{\lambda} \cong 1.5 \sqrt{\frac{\ell'}{\lambda}}$$

which implies that the plate is approximately as large as a paraboloidal dish whose gain equals that of the backfire antenna. For $\ell' = 2\lambda$, the gain is 19 dB, sidelobes are 12 dB down,

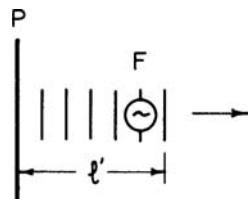


FIGURE 10-11 Backfire antenna (after Ehrenspeck¹⁶ 1959)

and bandwidth is at least 20 percent. Easier and cheaper to build than a paraboloidal dish, the backfire antenna might be competitive for gains up to 25 dB (provided sidelobes do not have to be very low); it is superior to ordinary end-fire antennas as long as a low silhouette is not a requirement.

Design for Minimum Beamwidth

The approximate relation between E -plane beamwidth BW_E , H -plane beamwidth BW_H , and gain of a surface-wave antenna is

$$G \cong \frac{30,000}{BW_E BW_H}$$

The half-power beamwidth of a maximum-gain design (Eq. 10-16) can be made 10 percent narrower by increasing the direct-feed radiation, which causes it to interfere destructively with $T(\theta)$ at a smaller angle θ than in Figure 10-7. This is done by starting the feed taper with λ/λ_z less than 1.25. The narrowest beamwidths reported in the literature are plotted in Figure 10-10 as the solid-line margin of the shaded region marked “Beamwidth.” (Usually E -plane patterns are slightly narrower; H -plane patterns, slightly wider.) Sidelobes are about 1 dB higher than in the maximum-gain case.

Design for Minimum Sidelobe Level

The fairly high sidelobes of the maximum-gain design can be reduced by increasing λ/λ_i to 1.35 at the start of the feed taper, letting $\lambda/\lambda_i \cong 1.2$ at the end of the feed taper, and continuing to taper over a large fraction of the total antenna length (to P in Figure 16-6*b*). In the constant cross-section region between P and T , λ/λ_i is chosen to lie on the lower margin of the shaded region in Figure 10-9. If P is two-thirds of the distance from F to T , the sidelobes of a 6λ -long antenna are down at least 18 dB in the H plane and 17 or 21 dB in the E plane of a dielectric rod or Yagi-Uda, respectively, at the cost of a 1.5-dB drop in gain and a 10 percent rise in beamwidth.^{17,18} As in the case of the optimum-gain design, the sidelobe level decreases monotonically with increasing antenna length.

The backlobe is suppressed by increasing the end-fire directivity of the feed pattern. A 25- to 30-dB level can be achieved without resorting to feeds that are much larger than the antenna cross section.

By tapering more sharply near F , then flattening out gradually to a value of λ/λ_z close to 1 at T , the H -plane sidelobe level can be reduced to 20 dB.¹⁸ Gain and beamwidth then lie on the dashed margins of the shaded bands in Figure 10-10. A further reduction to 30 dB in the H -plane sidelobe level is achieved by placing parasitic side rows on either side of the center array¹⁹ (see Figure 10-12). The side-row length ℓ' is slightly longer than 0.5ℓ . λ/λ_z is the same as on the center array (use Eq. 10-14*b*), and the spacing d is optimum when the side rows lie just beyond the 11-dB contour of the surface wave on the center array; or, using Figure 10-3*b* and Eq. 10-14*b*,

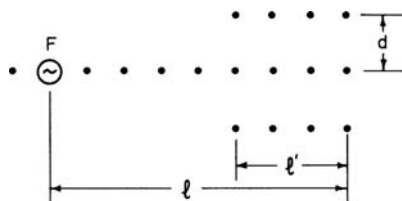


FIGURE 10-12 Parasitic side rows for sidelobe suppression in one plane (top view; dots indicate dipoles or monopoles above a ground plane).

$$\frac{d}{\lambda} \cong 0.25 \sqrt{\frac{\ell}{\lambda}}$$

The gain is the same as for maximum-gain design (Eq. 10-15). To reduce sidelobes in all planes, parasitic side rows would have to be placed all around the center array.

Design for Broad Pattern Bandwidth²⁰

The ± 15 percent bandwidth of most surface-wave antennas can be extended to as much as ± 33 percent (2:1) at the cost of a 2-dB gain decrease from optimum at mid-frequency. At the feed, the mid-frequency value of λ/λ_z should be 1.4. A uniform taper extends from the feed to the termination, as shown in Figure 10-1a. At the low-frequency end of the band, gain and beamwidth lie close to the solid margins of the shaded strips in Figure 10-10—at the high-frequency end, close to the dashed margins. Thus gain is constant within ± 1.5 dB and beamwidth within ± 15 percent. On a structure 13.5λ long at mid-frequency, sidelobes were down at least 11 dB throughout the band.²⁰

The bandwidth figures just quoted apply to the polyrod ($\epsilon = 2.56$) and to all surface-wave antennas whose dispersion curve (variation of λ/λ_i with frequency) resembles that of the polyrod. Dispersion increases with increasing dielectric constant; dielectric rods with $\epsilon < 2.56$ are therefore more broadband and rods with $\epsilon > 2.56$ less broadband than polyrods. The dispersion of artificial dielectrics varies; corrugated surfaces are comparable with polyrods, and Yagi-Udas are considerably more dispersive (precise bandwidth figures are not known).

The input impedance of most surface-wave antennas is slightly capacitive and changes slowly with frequency. A tapered dielectric section in waveguide (see Figures 10-1a and 10-6b) provides good matching over a 2:1 bandwidth if 1.5λ long at mid-frequency and over a smaller bandwidth if shorter. To minimize the voltage-standing-wave-ratio (VSWR), an inductive iris is placed close to the insertion point of the rod. Equivalent techniques apply in the case of antennas other than dielectric rods and of input transmission lines other than waveguide.

Feeds

Design principles of feeds and feed tapers for optimum gain, sidelobes, and bandwidth have already been stated. Specific structures will now be described.

In Figure 10-6a, the feeder and reflector monopoles are spaced 0.2λ apart, $h_r = 0.23\lambda$, and $h_f = 0.21\lambda$.¹³ These figures apply to $\delta = 0.048\lambda$; if the elements are thicker, they must be made slightly shorter. The backlobe is between 12 and 15 dB down. By replacing the reflector monopole with a semicircular plate (radius = h_r) or a small corner or paraboloidal reflector (height = h_r), we increase the gain of the feed by at least 1.5 dB and, therefore, the total antenna gain by approximately the same amount. The sidelobes are decreased by at least 2 dB; the backlobe is 25 to 30 dB down.

To couple from the dominant mode in rectangular waveguide, we replace the feeder monopole by a metal half ring²¹ (see Figure 10-13a); the excitation efficiency of this feed, with semicircular reflector plate, is around 80 percent. Another, slightly less efficient but simpler waveguide feed uses two or more slots to couple into the surface-wave structure. The waveguide should be dielectric-loaded so that its phase velocity approximates that in the surface-wave structure, and the slots are spaced about a quarter wavelength apart in the dielectric. Instead of cutting slots, the entire broad face of a dielectric-filled waveguide can be opened (see Figure 10-13b); by tapering the height of the waveguide over the appropriate length, feed and feed taper can be combined in a single design. In low-power applications, striplines can be slot- or dipole-coupled to the antenna.²² In the absence of the ground plane, we use a center-fed dipole (coaxial input with built-in balun, Figure 10-13c) or a folded dipole (usually preferred with two-wire input).

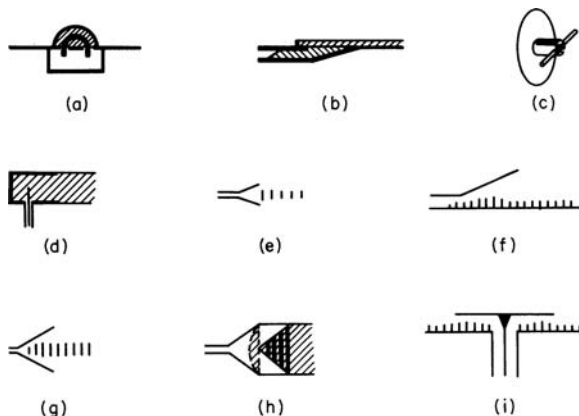


FIGURE 10-13 Surface-wave-antenna feeds: (a) Semicircular wire coupled to waveguide (after DuHamel and Duncan²¹ © IRE 1958); (b) Dielectric-filled waveguide with an open broad wall; (c) Dipole and reflector feed; (d) Circular waveguide cap; (e) and (f) Horn exciters; (g) Parallel-wire (distributed) feed (after Reynolds²⁴ © IRE 1957); (h) Flared transition with lossy strips (with collimating lens shown in horn aperture) (after Stephenson and Walter²⁸ © IRE 1955); (i) Coaxial excitation of disk

The coaxially fed metal cap (see Figure 10-13d), popular as a dielectric-rod feed or a cigar-antenna feed at ultrahigh frequencies (UHF) because of good efficiency combined with mechanical strength, forms the transition between feed types *a* and *b* of Figure 10-6. Type *b* can handle much higher power than type *a*. In Figure 10-6b the waveguide can be rectangular (with TE_{10} mode) or circular (TE_{11} mode). A variant of this form, suitable for Yagi-Uda or cigar-antenna excitation, is shown in Figure 10-13a. If the horn aperture is considerably larger than the antenna cross section (see Figure 10-13f), the surface-wave structure should start well inside, and it is desirable to have the horn extend just beyond the end of the feed taper, as indicated. (Theoretical work shows that for their size, horn apertures are not very efficient surface-wave exciters,²³ but if the direct-feed radiation of the horn is well collimated—the design can be based on Chapter 14—a large total antenna gain will be produced.)

Patterns that closely approximate the $\sin \xi/\xi$ shape have been obtained by flaring a parallel wire into a 60° V (see Figure 10-13g)²⁴ or by running an unflared parallel wire (loaded with transverse wire stubs to reduce phase velocity) alongside the antenna;²⁵ both structures extend over a third of the antenna length. Feeds of this type are called *distributed*. In the hope of obtaining good phase and amplitude control over the entire antenna aperture, feeds have been studied that couple continuously from one end of the antenna to the other, for example, a waveguide slot-coupled to a dielectric slab.^{26,27} These designs have not been very successful thus far, principally because energy flow along two such closely coupled waveguides is difficult to control. If precise aperture illumination is desired, we must abandon surface-wave antennas and turn to the end-fire structures among the surface-wave-excited arrays of discrete elements.

Variants of Figure 10-6a and *b* are also used to excite surface-wave area sources. Rectangular dielectric or corrugated surfaces are fed by placing a slotted waveguide alongside, or a hog horn, or a horn with correcting lens. The last is frequently plagued by transverse standing waves, which produce sidelobes in azimuth ($y\phi$ plane). To suppress these, a flared transition (1λ long at mid-frequency) with transverse conducting (Aquadag) strips

can be used (see Figure 10-13*h*).²⁸ It keeps sidelobes below 20 dB over a frequency range of at least 2:1 at a cost of 1 or 2 dB because of the lossy strips.

Radial-cylindrical area sources (disks, see Figure 10-2*d*) can be center-fed with a simple monopole feeder; its efficiency of excitation turns out to be higher than with line sources. A higher-gain is shown in Figure 10-13*i*, in which the coaxial input is excited in the TEM mode for vertical polarization on the disk and in the TE_{01} mode for horizontal polarization. A completely flush installation could be achieved by coupling through two or more annular slots. All the arrangements previously described for line sources can be carried over, and the optimum design principles for feed and feed taper still apply.

Arrays of Line Sources

Yagi-Uda, dielectric rods, and other surface-wave line sources can be arranged in a plane or volume array. The surface-wave illumination in the terminal plane of a plane array is scalloped (see Figure 10-14*a*), the overlapping dotted circles representing the 25-dB contours of the transverse-field decay. If identical end-fire elements and negligible interaction between them are assumed, the array gain G_a is the sum of the individual gains G . By letting n be the number of elements and using Eq. 10-15,

$$G_a = nG \cong 10n \frac{\ell}{\lambda} \quad (10-18)$$

Interaction between adjacent elements is found to be negligible if the crossover is at about the 10-dB contour; the minimum separation d is therefore (see Figure 10-3*b* and Eq. 10-14*b*)

$$\frac{d}{\lambda} \cong 0.5 \sqrt{\frac{\ell}{\lambda}} \quad (10-19a)$$

The exact value depends on the polarization (coupling is less in the E than in the H plane) and on the direct coupling between feeds. Too close a spacing produces a loss in gain. An upper limit on the spacing is obtained by requiring that the first principle sidelobe of the array pattern be strongly attenuated through multiplication with the element pattern. The angle with end-fire of the array sidelobe must therefore be at least that of the first minimum in the element pattern, a criterion that corresponds to a crossover of about 22 dB, or a maximum separation

$$\frac{d}{\lambda} \cong \sqrt{\frac{\ell}{\lambda}} \quad (10-19b)$$

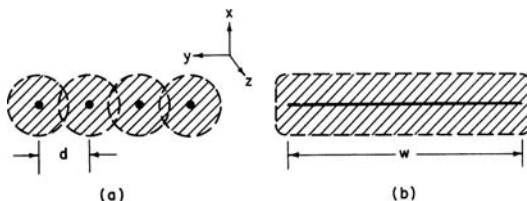


FIGURE 10-14 Two-dimensional surface-wave antennas (transverse cross sections): (a) array of line sources and (b) area source

The principle of pattern multiplication, which has just been invoked, is useful only for a first design, since the element patterns are somewhat distorted by the presence of the other elements. The optimum spacing, which lies somewhere between the d/λ values given by Eqs. 10-19, must therefore be found empirically.

The beamwidth BW_a in the plane of the array (azimuth) is approximately

$$BW_a \cong \frac{65\lambda}{nd} \quad (10-20)$$

Sidelobes can be reduced by tapering the amplitude at the feeds. In the elevation plane, beamwidth and sidelobes are controlled by the element pattern. The bandwidth is less than for individual elements, probably about ± 5 percent.

Low-silhouette requirements often force the choice of a horizontal plane array of end-fire elements in place of a vertical cylindrical reflector (or a rectangular mattress array). The trade-off between them, if equal gain and sidelobes and equal width in the y direction (see Figure 10-14a) are assumed, is

$$\frac{\ell}{\lambda} \cong \left(\frac{h}{\lambda} \right)^2 \quad (10-21)$$

(cf. Eq. 10-17), where ℓ is the length of the end-fire elements and h the heights (in the x direction) of the equivalent reflector or broadside array. Because of space limitations, the line sources cannot always be equal in length. Equation 10-18 is then inapplicable, but the spacing between elements is still controlled by Eqs. 10-19.

A volume array of end-fire elements occupies almost as large an area in the transverse (xy) plane as the dish it replaces and has advantages in terms of transportability, ease of erection, and minimum weight needed for mechanical strength against certain kinds of stresses. Spillover is more easily controlled than with a conventional dish, but sidelobe control becomes more difficult and the bandwidth is much narrower. In the design of these arrays, the number of end-fire elements can be reduced by increasing their length; according to Eq. 10-18, $n = G\lambda/10\ell$. This relation clarifies the trade-off between an end-fire volume array and a broadside mattress array of dipoles: though we cannot increase the total gain by replacing the dipoles with higher-gain end-fire elements, we can reduce the number of elements (and thus of separate feeds) while keeping the gain constant.

Arrays of end-fire elements can be fed in cascade from a common transmission line (like slots in waveguide), but a corporate feed structure (branching transmission lines) produces better bandwidth.

Effect of Finite Ground Plane on the Radiation Pattern

Vertically polarized surface-wave line or area sources are often mounted on a ground plane whose finite length g (see Figure 10-15a) distorts the antenna pattern in two ways: by tilting the beam through an angle ψ away from end-fire and by broadening it to a half-power beamwidth BW' . The tilt is maximum when $g = 0$ (no ground plane in front of antenna):^{29,30}

$$\psi_{\max} \cong 60 \frac{\lambda}{\ell} \text{ degrees}$$

(which is a full half-power beamwidth BW of the unperturbed pattern); the beam then broadens so that the 3-dB level lies in the end-fire direction, and BW' is slightly over twice as wide as the unperturbed BW (see Figure 10-15b and c). The tilt can be reduced by slowing the

surface wave down from its infinite-ground-plane optimum, but this increases sidelobes and decreases gain. The gain, on the other hand, is optimized by speeding the wave up at the expense of increased tilt.³⁰

As g increases, the beam tilt decreases: when $g = \ell$, $\psi \cong 0.7\psi_{\max}$; when $g = 3\ell$, $\psi \cong 0.5\psi_{\max}$; when $g = 20\ell$, $\psi \cong 0.2\psi_{\max}$ —a very gradual approach to the infinite-ground-plane condition. The beamwidth approaches the unperturbed value more rapidly: when $g = \ell$, $BW' \cong 1.25BW$.²⁹ Modifying the surface-wave velocity again has an opposite effect on beam tilt and gain.

If no beam tilting or loss in gain can be tolerated, flush mounting must be abandoned in favor of a full-size structure (pod-mounted to reduce drag, if necessary). A less drastic remedy is to reduce the tilt angle, at a small cost in gain, by bending the antenna and ground plane into a cylindrical or spherical cap.³¹⁻³³ The curvature produces an attenuation in the surface wave, which can be enhanced by tapering the antenna along its entire length (see Figure 10-15*d*). If the total attenuation is such that about 50 percent of the power has leaked off before the surface wave reaches the termination, the tilt angle will be reduced to near zero no matter how short the ground plane.³⁰

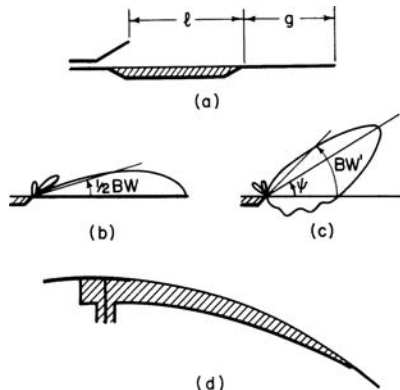


FIGURE 10-15 Effect of finite ground plane on the radiation pattern of an end-fire antenna: (a) Finite ground plane; (b) Pattern on infinite ground plane; (c) Pattern in the absence of ground plane; (d) Antenna on a curved ground plane

10.4 SURFACE-WAVE ANTENNAS: SPECIFIC STRUCTURES

Numerical values of λ/λ_z and design details for specific structures are presented in this section. The λ/λ_z curves are displayed for several values of relative dielectric constant ϵ , ranging from 2.56 (polystyrene) to 165 (a calcium titanate ceramic). Hard, low-loss, high-temperature materials, for example, fused quartz ($\epsilon \cong 3.7$), pyroceram ($\epsilon \cong 5.6$), and alite (an aluminum oxide, very durable, $\epsilon \cong 8.25$), are of special interest.

The dielectric-loss tangent δ , which is temperature-dependent, produces a plane-wave attenuation

$$\alpha\lambda = \sqrt{\epsilon} \tan \delta$$

in an infinite medium, but since a large fraction of the surface wave propagates outside the medium, the surface-wave attenuation $\alpha_z\lambda$ is smaller. For $\lambda/\lambda_z = 1.2$, $\alpha_z \cong 0.7\alpha$; for $\lambda/\lambda_z = 1.03$, $\alpha_z \cong 0.1\alpha$.³⁴ As long as δ is smaller than 10^{-2} , the loss will therefore be less than 0.1 dB per wavelength. Data on long Yagi-Uda³⁵ indicate that, for λ/λ_z near 1, loss in artificial dielectrics is likewise negligible, provided all joints are firmly press-fit or soldered and the metal is of high conductivity. As the corrugations or dipole elements approach resonance, λ/λ_z increases and ohmic losses rise sharply. A typical figure for loss in a 6λ-long dielectric rod or Yagi-Uda is 0.5 dB. (Since the maximum-gain curve in Figure 10-10 is based on experiment, the effect of ohmic loss is included.) Surface-wave losses can be measured by the resonator method.³⁶

Because of low frequencies, the metal elements of artificial dielectrics can be constructed of slender poles and chicken wire; they weigh far less than solid dielectrics. Large ground structures at UHF or below are therefore invariably artificial dielectrics. In the microwave range, the choice of medium depends on mechanical strength, temperature behavior, erosion resistance, and cost; ordinary dielectrics, some of which can be cheaply molded into cavities recessed in the skin, are usually preferred for flush airplane or missile installation.

Dielectric Rod^{8,17,20,21,34,37-42}

Dielectric rod is termed *polyrod* and *ferrod* when made of polystyrene and ferrite material respectively. See Figure 10-16a for phase velocity as a function of rod diameter. The higher the dielectric constant, the thinner the rod (for a given gain) and the lighter therefore the weight; however, the larger the dispersion and therefore the narrower the bandwidth.

Modes higher than the fundamental HE_{11} cannot propagate provided $d/\lambda < 0.626 / \sqrt{\epsilon}$. The first two higher modes, TE_{01} and TM_{01} , produce a null in the end-fire direction.

Cross sections other than circular are often useful. To a good approximation,³⁴ the phase velocity depends only on the cross-section area A , and Figure 10-16a therefore remains valid if d is replaced by $1.13 \sqrt{A}$. Sliced in half along the plane of electric symmetry (see Figure 10-4c)—which does not affect the phase velocity—a circular or rectangular rod can be placed on a ground plane (*dielectric-image line*, Figure 10-17a and b); when recessed for flush mounting, this structure is called the *dielectric channel guide* (see Figure 10-17c). Another variant, the dielectric tube,⁴² has also been studied but appears to have no advantages over ordinary rods.

Mallach³⁸ found that for optimum design the rod diameter at the feed end of a linearly tapered dielectric rod should be $d_{\max}/\lambda \cong [\pi(\epsilon - 1)]^{-1/2}$, and at the termination $d_{\min}/\lambda \cong [2.5\pi(\epsilon - 1)]^{-1/2}$. In the range $2.5 < \epsilon < 20$, this rule corresponds to letting λ/λ_z at the feed be roughly equal to 1.1 and at the termination to 1.0. The resulting patterns are often adequate but can be improved by using the methods described in Section 10.3.

A rough approximation to the rod pattern, due to Zinke,³⁸ consists in multiplying the $\sin \xi/\xi$ pattern shown in Figure 10-7 by the factor $\cos [(\pi d/\lambda) \sin \theta]$ (but the predicted sharp nulls do not in fact exist).

Figure 10-16b shows details of a conventional feed that has proved itself in practice. Other feeds are illustrated in Figures 10-6 and 10-13. In the case of waveguide feeds, it has

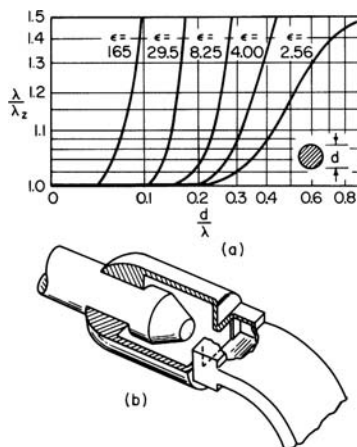


FIGURE 10-16 Dielectric rod: (a) Relative phase velocity (see the literature³⁷⁻³⁹; in case of a noncircular cross section, replace d/λ by $1.13 \sqrt{A}/\lambda$); (b) Conventional feed (after Southworth⁴³ © G. C. Southworth and D. Van Nostrand Company, Inc. 1950)

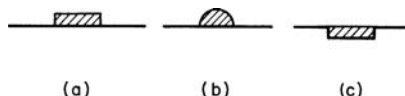


FIGURE 10-17 Dielectric image lines and channel guide (a) and (b) image lines (identical λ/λ_z if cross section, areas are the same) and (c) channel guide

been found desirable,¹⁷ both for snug fitting and for sidelobe suppression, to wrap dielectric tape around the feed point or, equivalently, to let the diameter of the waveguide feed be slightly smaller than that of the dielectric rod at the feed point, as shown in Figure 10-6b. If $\epsilon > 8$, the rod can be fed over a narrow band by direct insertion in the narrow side of a waveguide, or in a cavity.⁴⁴

Dielectric Channel Guide^{28,45,46} (Figure 10-17c)

Two modes are of interest: a vertically polarized one that is a deformation of the HE_{11} rod mode and a horizontally polarized one. The phase velocity of the vertically polarized mode is very slightly faster in Figure 10-17c than in Figure 10-17a or b. The mode may be slightly leaky (small attenuation in the axial direction), but this is not known with certainty. The structure can be viewed as a channel of the type shown in Figure 10-2b, but with most of the sidewalls removed. The phase velocity is then determined as follows: we rewrite Eq. 10-5b in the form

$$\frac{\lambda}{\lambda_z} = \sqrt{\left(\frac{\lambda}{\lambda_z'}\right)^2 - \left(\frac{\lambda}{2\omega}\right)^2}$$

where λ/λ_z' is the relative phase velocity on an infinitely wide slab-on-metal, as given in Figure 10-19c. Removal of the sidewalls above the dielectric-air interface appears to have negligible effect on the phase velocity²⁸ and, at least as long as $\lambda/\lambda_s > 1$ (ω is not too narrow), introduces little if any leakage attenuation.

The horizontally polarized mode is the lowest-order TE mode in the channel of Figure 10-2b (slightly perturbed by the removal of most of the sidewalls), which in turn is identical with the lowest-order TE mode on a dielectric slab-on-metal (solid curves, see Figure 10-19d).

The dielectric-filled channel also supports leaky waves, and care must be taken not to choose parameters that allow these to be excited.

Yagi-Uda^{13,19,47-62}

These are sometimes referred to as *ladder* arrays. Figure 10-18 shows that the phase velocity is controlled by adjusting the spacing, height, and diameter of the monopole elements.

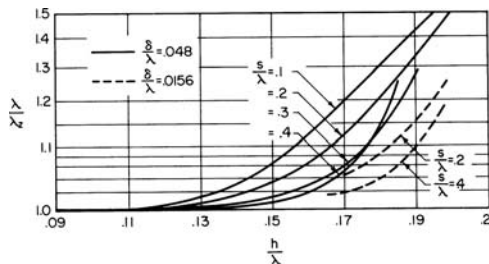


FIGURE 10-18 Relative phase velocity on Yagi-Uda antenna (after Ehrenspeck and Poehier¹³ © IRE 1959). For the meaning of parameters, see Figure 10-6a.

The dispersion is more pronounced than on a dielectric rod, and the best-reported bandwidth (within which the gain drops no more than 3 dB) is only ± 10 percent; this can perhaps be improved by application of the broadband design method described in Section 10.3. The bandwidth of an array of Yagis is at least ± 5 percent.

The curves in Figure 10-18 apply to the row of monopoles shown in Figure 10-6*a* and equally to a row of dipoles (element height $2h$). A center boom for mounting the dipole elements has negligible effect on the phase velocity provided the total dipole span $2h$ includes the boom diameter. Yagi-Udas can be photoetched on a copper-clad dielectric sheet;^{50,51} since the dielectric slows down the wave, the element height needed to produce a given phase velocity is less than indicated in Figure 10-18.

Dielectric Sheets and Panels^{4,7,31-33,63-68}

Because of the imaging properties of vertically polarized waves, the lowest-order TM mode on a dielectric sheet-on-metal is identical with that on a panel of twice the sheet thickness (see Figure 10-19*a-c*). This wave has no cutoff frequency; its properties are discussed in Section 10.2, and Figure 10-4 shows the field configuration.

The lowest-order horizontally polarized wave on a dielectric sheet-on-metal produces a null in the end-fire direction and has a cutoff (solid curves, Figure 10-19*d*). On a panel, on the other hand, the lowest-order horizontally polarized wave produces a maximum in the end-fire direction and has no cutoff (dashed curves, see Figure 10-19*d*). (The next higher TE mode on the panel is identical with the lowest TE mode on the sheet, but because of the end-fire null it is not a useful mode.)

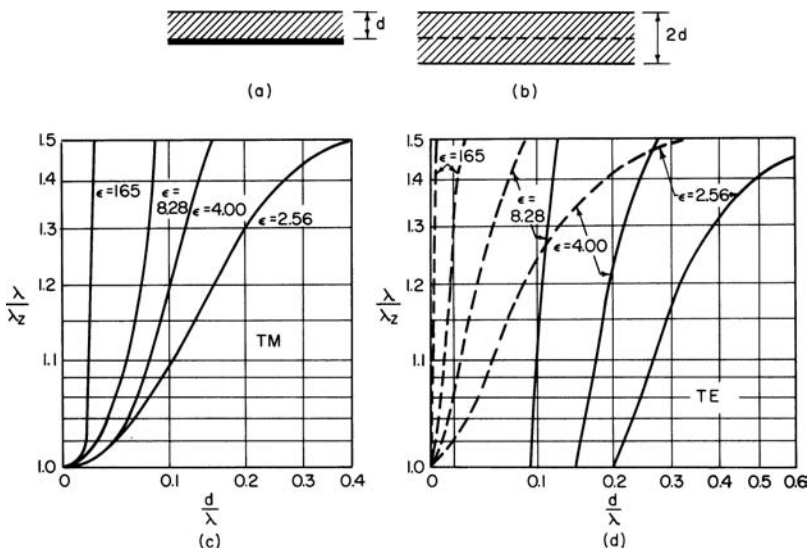


FIGURE 10-19 Dielectric sheets and panels: (a) Sheet-on-metal; (b) Panel; (c) Lowest-order TM mode on sheet or panel; (d) Lowest-order TE mode on sheet (solid curves) and panel (dashed curves)

In designing tapered sheets on metal, a flush air-dielectric interface can be maintained by contouring the depth of ground plane. Good contact between dielectric and ground plane is needed, since air spaces affect the phase velocity.⁷

Circular dielectric disks⁶⁶ and spherical caps^{32,33} are useful as omnidirectional beacon antennas. For design principles, see Section 10.3 (paragraphs on the effect of finite ground plane on the radiation pattern and on beam-shaping techniques). Figure 10-19c and d are still applicable, since λ/λ_z is only slightly affected by a decreasing radius of curvature. The leakage attenuation has been calculated by Elliott.³² Patterns provide good null filling to 45° off end-fire and have a bandwidth of ± 8 percent.³³

Circular (or cross) polarization can be produced if the surface-wave structure supports a TM and a TE wave with identical phase velocities. One example is the two-layer dielectric sheet,^{7,67} illustrated in Figure 10-20a for the case in which the lower layer is air. Given λ/λ_z , Figure 10-20b prescribes the layer thickness (with $\epsilon = \epsilon_1/\epsilon_0$ as a parameter). Another, structurally superior, example is the simple-layer dielectric in which longitudinal metal vanes that short out the vertically polarized mode while leaving the horizontally polarized one unperturbed are embedded (see Figure 10-20c). The TM wave therefore “sees” a thinner sheet than the TE wave, and for a prescribed λ/λ_z Figure 10-20b gives the values (in first approximation) of the vane height and total layer thickness. For details, see Hansen.⁶⁸

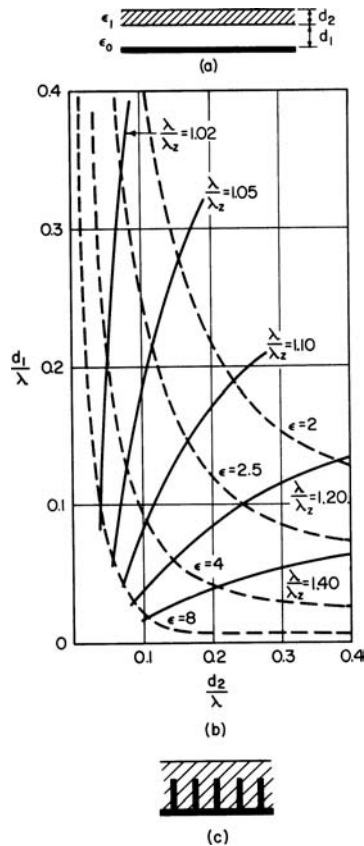


FIGURE 10-20 Circular (or cross) polarization on dielectric sheets: (a) A simple double-layer structure; (b) Parameter combinations (with $\epsilon = \epsilon_1/\epsilon_0$) that allow (a) to support a circularly polarized surface wave with λ/λ_z between 1.02 and 1.40 (after Plummer and Hansen⁶⁷ © IEE 1957); (c) Longitudinal metal vanes embedded in dielectric (after Hansen⁶⁸ © IRE 1957)

10.5 LONG-WIRE ANTENNAS⁶⁹⁻⁷⁵

Introduction

Long-wire antennas were in active development during the height of the high-frequency era, before 1940, and many different forms evolved. Since then development has been quiescent, the only surviving extant form being the horizontal rhombic antenna. Its continued popularity stems from its simple construction, low cost, and wideband capabilities. The rhombic antenna is uniquely adapted to the propagation of high frequencies via the ionosphere. After 1950, versions of the log periodic antenna have replaced the rhombic for hf communications. (See Chapter 13 for details of log periodic designs.)

The configurations comprising many obsolete long-wire antennas, including the vertical Franklin arrays and the resonant horizontal types, are now history and can be found in the literature.

Some space in this chapter will be devoted to the Beverage antenna, which has desirable applications to the directive reception of low and very low frequencies.

Single Long Wire as a Unit Radiator

Radiation Patterns

The following basic facts about the radiation patterns of long wires isolated in free space with idealized natural current distributions are useful for reference:

1. There is a lobe in the radiation pattern for each half wavelength of wire length. Each lobe is a cone of radiation centered on the wire.
2. With respect to the middle of the wire as center of the polar radiation pattern, half of the lobes are tilted forward and half of them backward. If the wire is an odd number of half wavelengths long, there is an odd number of lobes; so one of them will be normal to the wire.
3. The direction of the electric field reverses in each successive lobe. This is analogous to the reversals of phase of the currents in successive half-wavelength portions of the wire.
4. Between successive lobes are regions of little or no radiation, called nulls or zeros. Practically, these are minima because the field strengths at these angles never actually go to zero. These zeros are related analogously to the current zeros in the standing-wave current distribution or to the phase reversals in a traveling-wave distribution.
5. The angles of the *zeros* are symmetrically distributed about the plane normal to the middle of the wire.
6. For a given electrical length, the angles of the *zeros* are the same for both standing- and traveling-wave current distributions.
7. In the first quadrant the angles of the maxima for both current distributions are virtually coincident. The only exception applies to the direction of the main (first) lobe for wire lengths less than three wavelengths, as may be seen in Figure 10-21.
8. The largest lobe in the radiation pattern is the one forming the smallest angle (θ_1) to the wire. For the traveling-wave case, this lobe is in the direction of current flow only. For the standing-wave case, the radiation pattern is symmetrical with respect to the middle of the wire and so has a complementary major lobe toward the other end of the wire.
9. With a standing-wave current distribution an integral number of half wavelengths long, the envelope of the polar plot of the field-strength distribution pattern is a straight line parallel to the wire or, more correctly, a concentric cylinder. For a nonintegral number of half wavelengths, the middle lobes are further decreased in amplitude. The maximum depression of inner lobes occurs when the current distribution consists of an odd integral number of quarter wavelengths (see Figure 10-22).
10. With traveling waves, the field-strength pattern has lobes of diminishing amplitude, the smallest being in the direction opposite that of current flow.
11. Typical deviations from idealized current distributions modify the relative lobe amplitudes slightly and fill in the zeros slightly but do not affect the *angles* of the maxima or the zeros in the radiation pattern.
12. The ratios of the amplitudes of successive lobes are higher for a traveling-wave system than for the equivalent standing-wave system (compare Figures 10-23 and 10-24).

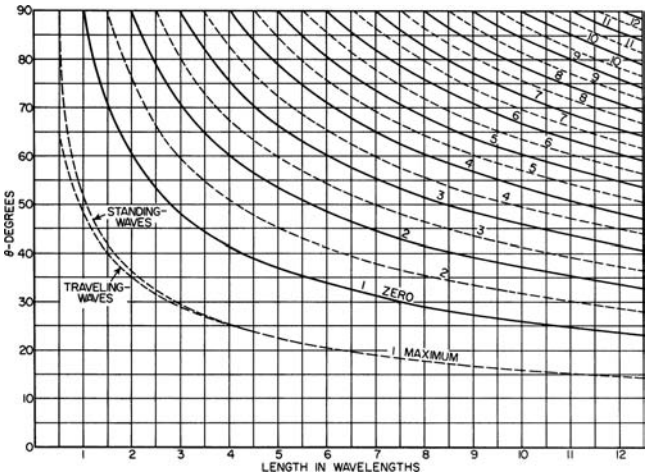


FIGURE 10-21 Angles of maxima and zeros in the radiation patterns for isolated long wires with natural current distribution

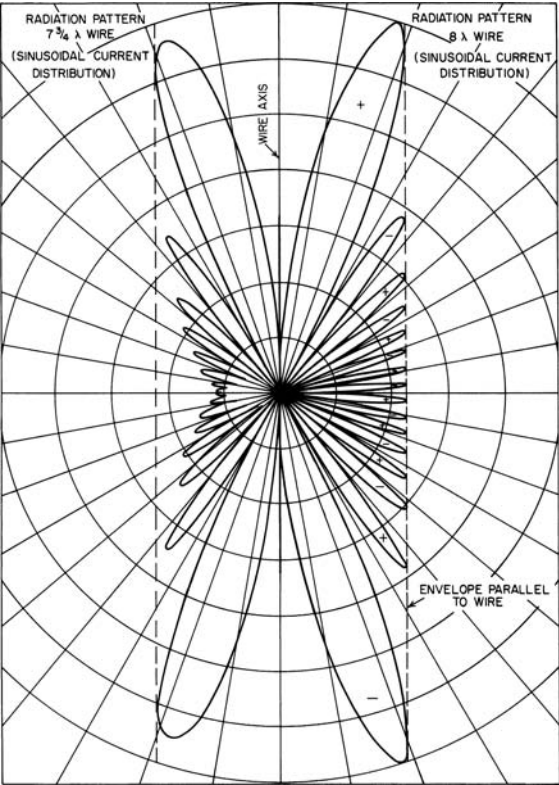


FIGURE 10-22 Comparison between the polar radiation patterns of 7.75λ - and an 8λ -long wire with standing-wave current distribution, in terms of relative field strength

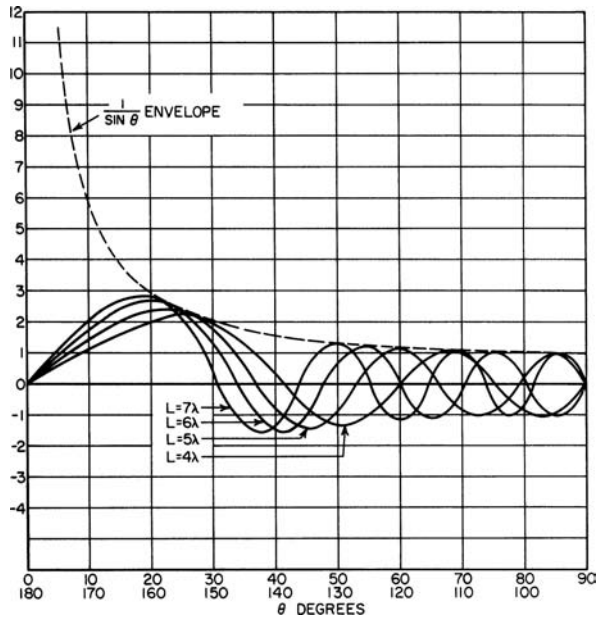


FIGURE 10-23 Field-strength radiation patterns for standing-wave current distributions

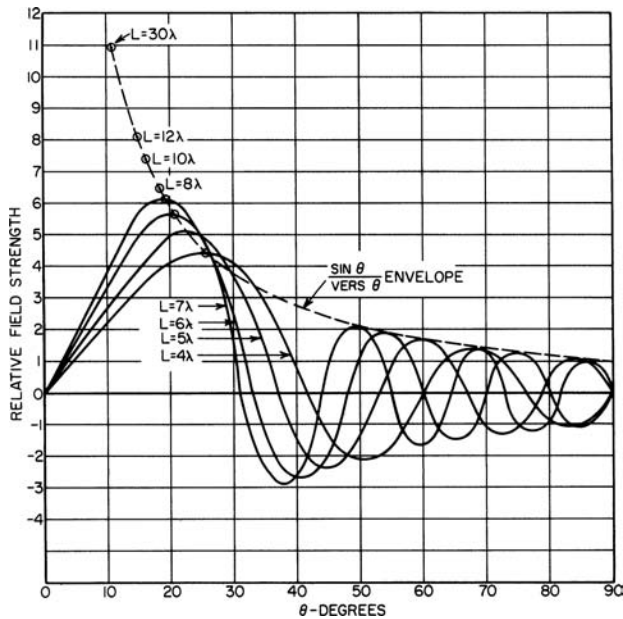


FIGURE 10-24 Field-strength radiation patterns for traveling-wave current distributions

Comparisons between standing-wave and traveling-wave field-strength patterns can be seen in Figures 10-23 and 10-24.

Horizontal Rhombic Antenna

This antenna is constructed as an elevated diamond with sides from two to many wavelengths long. It is used for transmission and reception of high-frequency waves propagated via the ionosphere. When terminated with a resistance equal to its characteristic impedance at its forward apex, it functions as a traveling-wave antenna; the termination suppresses reflections of transmitted power and absorbs signals from the contrary direction.

Figure 10-25 illustrates optimum electrical parameters and the limits of frequency ratios within which performance is not unreasonably compromised. Note that with less than two wavelengths per leg, the effectiveness of a rhombic antenna is generally regarded as unacceptable because its gain is too low, and in transmission a high percentage of the input power must be dissipated in a terminal resistor. With legs of seven and eight wavelengths, the terminal loss is down to 18 to 24 percent because of the radiation efficiency of such structures.

A resistance-terminated rhombic becomes essentially a transmission line, but this analogy is compromised by the expansion of the line toward the sides and its reconvergence to the forward apex. To improve the uniformity of distributed line constants and to reduce the rhombic's characteristic impedance to convenient values such as $600\ \Omega$, three wires expand from and to each apex on each side. They are at their greatest spread at the sides. The side angles disturb the propagation of currents enough to cause small reflections, but these usually are negligible.

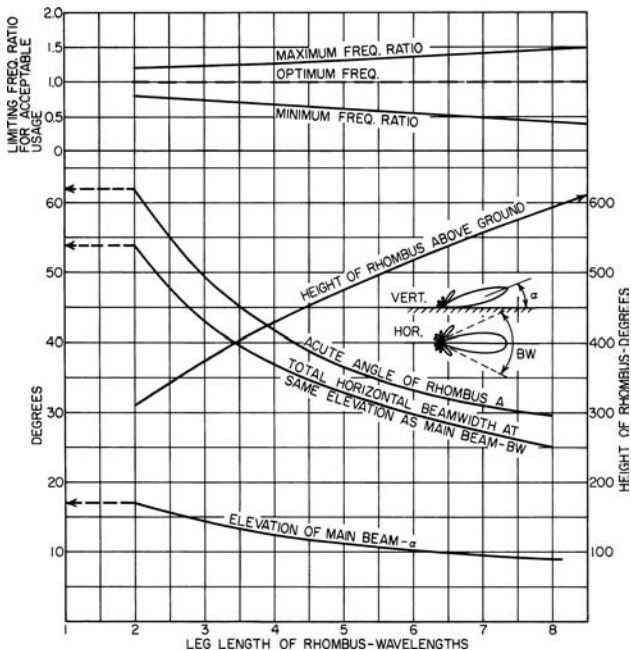


FIGURE 10-25 Optimum design parameters for high-frequency horizontal rhombic antennas

REFERENCES

1. F. J. Zucker and J. A. Strom, "Experimental Resolution of Surface-Wave Antenna Radiation into Feed and Terminal Patterns," *IEEE Trans. Antennas Propagat.*, vol. AP-18 (May 1970): 420-422.
2. F. J. Zucker, "Surface-Wave Antennas," Chap. 21, R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, part 2 (New York: McGraw-Hill, 1968).
3. N. Marcuvitz (ed.), *Waveguide Handbook, MIT Rad. Lab. ser.*, vol. 10 (New York: McGraw-Hill, 1951).
4. S. S. Attwood, "Surface Wave Propagation over a Coated Plane Conductor," *J. App. Phys.*, vol. 22 (1951): 504.
5. J. H. Richmond and T. E. Tice, "Probes for Microwave Near-Field Measurements," *IRE Trans. Microwave Theory Tech.*, vol. MTT-3 (1955): 32.
6. H. Ehrenspeck, W. Gerbes, and F. J. Zucker, "Trapped Wave Antennas," *IRE Nat. Conv. Rec.*, part 1 (1954): 25.
7. J. B. Andersen, *Metallic and Dielectric Antennas* (Copenhagen: Polyteknisk Forlag, 1971).
8. J. Brown and J. O. Spector, "The Radiating Properties of End-Fire Aerials," *IEE Proc.*, part B, vol. 104 (1957): 25.
9. A. F. Kay, "Scattering of a Surface Wave by a Discontinuity in Reactance," *IRE Trans. Antennas Propagat.*, vol. AP-7 (1959): 22.
10. F. J. Zucker and J. A. Strom, "Experimental Resolution of Surface-Wave Antenna Radiation into Feed and Terminal Patterns," *IEEE Trans. Antennas Propagat.*, vol. AP-18 (May 1970): 420-422.
11. L. Felsen, "Field Solutions for a Class of Corrugated Wedge and Cone Surfaces," MRI Electrophys. Group Mem. 32, (Brooklyn, N.Y.: Polytechnic Institute of Brooklyn, 1957).
12. V. I. Talanov, "On Surface Electromagnetic Waves in Systems with Nonuniform Impedance," *Izv. VUZ MVO (Radiofiz.)*, vol. 2 (1959): 32. (Translated from the Russian by Morris P. Friedman, Inc., Air Force Cambridge Res. Cen. Rep. TN-59-768.)
13. H. W. Ehrenspeck and H. Poehler, "A New Method for Obtaining Maximum Gain from Yagi Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-7 (1959): 379.
14. H. W. Ehrenspeck and W. Kearns, unpublished experimental data, Air Force Cambridge Research Center, Bedford, Mass., 1959.
15. J. C. Simon and G. Weill, "A New Type of Endfire Antenna," *Annales de radioélectricité*, vol. 8 (1953): 183. (In French.)
16. H. W. Ehrenspeck, U.S. patent applied for, Air Force Cambridge Research Center, Bedford, Mass., 1959.
17. F. E. Boyd and D. H. Russell, "Dielectric-Guide Antennas for Aircraft," *Nav. Res. Lab. Rep.* 3814, 1951.
18. F. J. Zucker and W. Kearns, unpublished experimental data, Air Force Cambridge Research Center, Bedford, Mass., 1959.
19. H. W. Ehrenspeck and W. Kearns, "Two-Dimensional End-Fire Array with Increased Gain and Side Lobe Reduction," *IRE Wescon Conv. Rec.*, part 1 (1957): 217.
20. C. F. Parker and R. J. Anderson, "Constant Beamwidth Antennas," *IRE Nat. Conv. Rec.*, part 1 (1957): 87.
21. R. H. DuHamel and J. W. Duncan, "Launching Efficiency of Wires and Slots for a Dielectric Rod Waveguide," *IRE Trans. Microwave Theory Tech.*, vol. MTT-6 (1958): 277.
22. A. D. Frost, C. R. McGeogh, and C. R. Mingins, "The Excitation of Surface Waveguides and Radiating Slots by Strip-Circuit Transmission Lines," *IRE Trans. Microwave Theory Tech.*, vol. MTT-4 (1956): 218.
23. A. L. Cullen, "The Excitation of Plane Surface Waves," *IRE Proc.*, part IV, Monograph 93R, vol. 101 (1954): 225.

24. D. K. Reynolds, "Broadband Traveling Wave Antennas," *IRE Nat. Conv. Rec.*, part 1 (1957): 99.
25. D. K. Reynolds and R. A. Sigelmann, "Research on Traveling Wave Antennas," AFCRC-TR-59-160, final rep., Cont. AF19(604)-4052 (Seattle, WA: Seattle University, 1959).
26. W. L. Weeks, "Coupled Waveguide Excitation of Traveling Wave Antennas," *IRE Wescon Conv. Rec.*, part 1 (1957): 236.
27. R. R. Hodges, Jr., "Distributed Coupling to Surface Wave Antennas," Cont. AF33(616)-3220, (Urbana: University of Illinois, 1957).
28. B. T. Stephenson and C. H. Walter, "Endfire Slot Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-3 (1955): 81.
29. R. S. Elliott, "On the Theory of Corrugated Plane Surfaces," *IRE Trans. Antennas Propagat.*, vol. AP-2 (1954): 71.
30. R. A. Hurd, "End-Fire Arrays of Magnetic Line Sources Mounted on a Conducting Half-Plane," *Can. J. Phys.*, vol. 34 (1956): 370.
31. R. S. Elliott, "Azimuthal Surface Waves on Circular Cylinders," *J. Appl. Phys.*, vol. 26 (1955): 368.
32. R. S. Elliott, "Spherical Surface-Wave Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-4 (1956): 422.
33. R. E. Plummer, "Surface-Wave Beacon Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-6 (1958): 105.
34. S. P. Schlesinger and D. D. King, "Dielectric Image Lines," *IRE Trans. Microwave Theory Tech.*, vol. MTT-6 (1958): 29.
35. A. F. Kay, unpublished work under AFCRC Cont. AF19(604)-3476 (Somerville, MA: Technical Research Group, Inc., 1959).
36. E. H. Scheibe, B. G. King, and D. L. Van Zeeland, "Loss Measurements of Surface Wave Transmission Lines," *J. App. Phys.*, vol. 25 (1954): 790.
37. G. E. Mueller and W. A. Tyrrell, "Polyrod Antennas," *Bell Syst. Tech. J.*, vol. 26 (1947): 837.
38. D. G. Kiely, *Dielectric Aerials* (London: Methuen & Co., Ltd., 1953).
39. L. W. Mickey and G. G. Chadwick, "Closely Spaced High Dielectric Constant Polyrod Arrays," *IRE Nat. Conv. Rec.*, part I (1958): 213.
40. R. W. Watson and C. W. Horton, "The Radiation Patterns of Dielectric Rods—Experiment and Theory," *J. App. Phys.*, vol. 19 (1948): 661; "On the Calculation of Radiation Patterns of Dielectric Rods," *J. App. Phys.*, vol. 19 (1948): 836.
41. R. Chatterjee and S. K. Chatterjee, "Some Investigations on Dielectric Aerials," part I, *J. Ind. Inst. Sci.*, vol. 38 (1956): 93; part II, *J. Ind. Inst. Sci.*, vol. 39 (1957): 134.
42. R. E. Beam, "Wave Propagation in Dielectric Tubes," final rep., Cont. DA36-039-sc-5397, Northwestern University, Evanston, Ill., 1952.
43. G. C. Southworth, *Principles and Applications of Wave-Guide Transmission* (Princeton, N.J.: D. Van Nostrand Company, Inc., 1950).
44. F. Reggia, E. G. Spencer, R. D. Hatcher, and J. E. Tompkins, "Ferrod Radiator System," *IRE Proc.*, vol. 45 (1957): 344.
45. W. Rotman, "The Channel Guide Antenna," *Proc. Nat. Electron. Conf.*, vol. 5 (1949): 190.
46. D. K. Reynolds and W. S. Lucke, "Corrugated End-Fire Antennas," *Proc. Nat. Electron. Conf.*, vol. 6 (1950): 16.
47. J. O. Spector, "An Investigation of Periodic Rod Structures for Yagi Aerials," *IEE Proc.*, part B, vol. 105 (1958): 38.
48. A. D. Frost, "Surface Waves in Yagi Antennas and Dielectric Waveguides," final rep., Cont., AF19(604)-2154, AFCRC-TR-57-368 (Medford, MA: Tufts University, 1957).
49. D. L. Sengupta, "On the Phase Velocity of Wave Propagation along an Infinite Yagi Structure," *IRE Trans. Antennas Propagat.*, vol. AP-7 (1959): 234.
50. J. A. McDonough and R. G. Malech, "Recent Developments in the Study of Printed Antennas," *IRE Nat. Conv. Rec.* (1957): 173.
51. R. G. Malech, "Lightweight High-Gain Antenna," *IRE Nat. Conv. Rec.*, part I (1958): 193.

52. H. E. Green, "Design Data for Short and Medium Length Yagi-Uda Arrays," *Elec. Eng. Trans. Inst. Eng.* (Australia) (March 1966): 1–8.
53. R. J. Mailloux, "The Long Yagi-Uda Array," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (March 1966): 128–137.
54. P. P. Viezbicke, "Yagi Antenna Design," NBS Tech. Note 688, National Bureau of Standards, U.S. Department of Commerce, December 1968.
55. G. A. Thiele, "Analysis of Yagi-Uda Type Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-17 (January 1969): 24–31.
56. D. K. Cheng and C. A. Chen, "Optimum Spacings for Yagi-Uda Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (September 1973): 615–623.
57. C. A. Chen and D. K. Cheng, "Optimum Element Lengths for Yagi-Uda Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (January 1975): 8–15.
58. N. K. Takla and L. C. Shen, "Bandwidth of a Yagi Array with Optimum Directivity," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (November 1977): 913–914.
59. J. L. Lawson, series of papers in *Ham Radio*: "Yagi Antenna Design: Performance Calculations" (January 1980): 22–27; "Yagi Antenna Design: Experiments Confirm Computer Analysis" (February 1980): 19–27; "Yagi Antenna Design: Performance of Multi-Element Simplistic Beams" (May 1980): 18–26; "Yagi Antenna Design: More Data on the Performance of Multi-Element Simplistic Beams" (June 1980): 33–40; "Yagi Antenna Design: Optimizing Performance" (July 1980): 18–31; "Yagi Antenna Design: Quads and Quagis" (September 1980): 37–45; "Yagi Antenna Design: Ground or Earth Effects" (October 1980): 29–37; "Yagi Antenna Design: Stacking" (November 1980): 22–34; "Yagi Antennas: Practical Designs" (December 1980): 30–41. See also corrections (September 1980): 67.
60. R. S. Elliott, *Antenna Theory and Design*, secs. 8.7 and 8.8 (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1981).
61. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, sec. 5.4. (New York: John Wiley & Sons, Inc., 1981).
62. C. A. Balanis, *Antenna Theory, Analysis and Design*, sec. 9.3.3 (New York: Harper & Row Publishers, Incorporated, 1982).
63. L. Hatkin, "Analysis of Propagating Modes in Dielectric Sheets," *IRE Proc.*, vol. 42 (1954): 1565.
64. F. E. Butterfield, "Dielectric Sheet Radiators," *IRE Trans. Antennas Propagat.*, vol. AP-3 (1954): 152.
65. R. L. Pease, "On the Propagation of Surface Waves over an Infinite Grounded Ferrite Slab," *IRE Trans. Antennas Propagat.*, vol. AP-6 (1958): 13.
66. E. M. T. Jones and R. A. Folsom, Jr., "A Note on the Circular Dielectric Disk Antenna," *IRE Proc.*, vol. 41 (1953): 798.
67. R. E. Plummer and R. C. Hansen, "Double-Slab Arbitrary-Polarization Surface-Wave Structure," *IEE Proc.*, part C, Monograph 238R (1957).
68. R. C. Hansen, "Single Slab Arbitrary Polarization Surface Wave Structure," *IRE Trans. Microwave Theory Tech.*, vol. MTT-5 (1957): 115.
69. E. A. Laport, *Radio Antenna Engineering* (New York: McGraw-Hill, 1952).
70. Donald Foster, "Radiation from Rhombic Antennas," *IRE Proc.*, vol. 25 (October 1937): 1327. A classic on rhombic-antenna theory and design.
71. E. A. Laport and A. C. Veldhuis, "Improved Antennas of the Rhombic Class," *RCA Rev.* (March 1960).
72. E. Bruce, A. C. Beck, and L. R. Lowry, "Horizontal Rhombic Antennas," *IRE Proc.*, vol. 23 (January 1935): 24.
73. A. E. Harper, *Rhombic Antenna Design* (New York: D. Van Nostrand Company, Inc., 1941).
74. F. A. Polkinghorn, "Single-Sideband MUSA Receiving System for Commercial Operation on Transatlantic Radio Telephone Circuits," *Bell Syst. Tech. J.* (April 1940): 306–335.
75. A. Baily, S. W. Dean, and W. T. Wintringham, "Receiving System for Long-Wave Transatlantic Telephony," *Bell Syst. Tech. J.* (April 1929): 309–367.

BIBLIOGRAPHY

- Agrawal, N. K. "A Radical Cylindrical Surface Wave Antenna," *J. Inst. Electron. Telecommun. Eng. (India)*, vol. 23 (July 1977): 448-450.
- Andersen, J. B. "Low and Medium Gain Microwave Antennas," in A. W. Rudge, K. Milne, A. D. Olver, and P. Knight (eds.), *The Handbook of Antenna Design*, vol. 1 (London: Peter Peregrinus Ltd., 1982), chap. 7.
- Blakey, J. R. "Calculation of Dielectric-Aerial-Radiation Patterns," *Electron. Lett.*, vol. 4 (1968): 46.
- Bojsen, J. H., H. Schjase-Jacobsen, E. Nilsson, and J. Bach Andersen. "Maximum Gain of Yagi-Uda Arrays," *Electron. Lett.*, vol. 7 (Sept. 9, 1971): 531-532.
- Chen, K. M. and R. W. P. King. "Dipole Antennas Coupled Electromagnetically to a Two-Wire Transmission Line," *IEEE Trans. Antennas Propagat.*, vol. AP-9 (1961): 405-432.
- Cho, S. H. and R. J. King. "Numerical Solution of Nonuniform Surface Wave Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-24 (July 1976): 483-490.
- Hall, P. S., B. Chambers, and P. A. McInnes. "Calculation of the Radiation Pattern, Gain and Input Impedance of a Yagi Antenna," *Electron. Lett.*, vol. 11 (June 1975): 282-283.
- Hessel, A. "General Characteristics of Traveling-Wave Antennas," in R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, part 2 (New York: McGraw-Hill, 1969), chap. 19.
- Hirasawa, K. "Optimum Gain of Reactively Loaded Yagi Antennas," *Trans. IECE of Japan*, vol. E63, no. 2 (Abstracts) (February 1980): 139.
- Hockham, G. A. and M. D. Byars. "A Broad Band, Compact Launcher for Surface Waves," *1969 European Microwave Conference* (New York: IEEE, 1970): 294-297.
- Inagaki, N., T. Sekiguchi, and M. Yamazawa. "Excitation of Surface Wave on Circular-Loop Array," *IEEE Trans. Antennas Propagat.*, vol. AP-19 (May 1971): 433-435.
- Kahn, W. K. "Double-Ended Backward-Wave Yagi Hybrid Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-29, no. 3 (1981): 530-532.
- Kajfez, D. "Nonlinear Optimization Reduces the Sidelobes of Yagi Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (September 1973): 714-715.
- . "Nonlinear Optimization Extends the Bandwidth of Yagi Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (March 1975): 287-289.
- Kamal, A. K., N. K. Agrawal, and S. C. Gupta. "Radiation Characteristics of Zig-Zag Antenna at 600 MHz," *J. Inst. Eng. (India) Electron. Telecommun. Eng. Div.*, vol. 59 (August 1978): 9-11.
- , S. C. Gupta, and R. A. Nair. "Dielectric Sphere-Loaded Radial Surface-Wave Antenna," *Int. J. Electron.* (Great Britain), vol. 44 (May 1978): 553-558.
- King, R. P. W. "Cylindrical Antennas and Arrays," in R. E. Collin and F. J. Zucker (eds.), *Antenna Theory*, part 2 (New York: McGraw-Hill, 1969), chap. 11.
- Kosta, S. P. "A Theory of Helical Yagi Antenna," *Nachrichtentech. Z. (Germany)*, vol. 25 (July 1972): 342-344.
- Morozov, B. N. and V. V. Chebyshev. "Design of a Corrugated Surface-Wave Antenna with Low Sidelobe Level," *Telecommun. Radio Eng.*, part 1 (United States), vol. 25 (April 1971): 26-28.
- Rojarayanot, B. and T. Sekiguchi. "Yagi-Uda Loop Antenna with Feeder System Composed of Two Elements Excited in Sequences," *Trans. IEEE of Japan*, vol. E61 (April 1978): 326-327.
- Rotman, W. and A. A. Oliner. "Asymmetrical Trough Waveguide Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-7 (1959): 153-162.
- Takata, N. and T. Sekiguchi. "Radiation of Surface Wave Antenna Consisting of Circular Loop Elements," *Trans. Inst. Electron. Commun. Eng. Japan, Sec. E*, vol. E60 (November 1977): 657-658.
- Tilton, E. P. *The Radio Amateur's VHF Manual* (Newington, Conn.: American Radio Relay League, Inc., 1972), p. 155.
- Tranquilla, J. M. and K. G. Balmain. "Resonance Phenomena on Yagi Arrays," *Can. Elec. Eng. J. (Canada)*, vol. 6 (April 1981): 9-13.
- Walter, C. H. *Traveling Wave Antennas* (New York: McGraw-Hill, 1965).

BIBLIOGRAPHY FOR LONGWIRE ANTENNAS

- Booth, C. F. and B. N. MacLarty. "The New High-Frequency Transmitting Station at Rugby," *J. IEEE (London)*, part B, vol. 103 (October 1955). Rhombic antenna layouts, two rhombic antennas on one set of masts, and feeder crosstalk measurements.
- Bruce, E. "Developments in Short-Wave Directive Antennas," *IRE Proc.*, vol. 19 (1931): 1406; *Bell Syst. Tech. J.*, vol. 10 (October 1951): 656. Evolution of the vertical-V antenna.
- Carter, P. S., C. W. Hansell, and N. E. Lindenblad. "Development of Directive Transmitting Antennas by RCA Communications," *IRE Proc.*, vol. 19 (October 1931): 1733. Evolution of standing-wave long-wire antenna development.
- CCIR Antenna Diagrams* (Geneva: International Telecommunications Union, 1954).
- Christiansen, W. N. "Directional Patterns for Rhombic Antennae," *AWA (Australia) Tech. Rev.*, vol. 7 (1946): 1. Study and pattern computations for horizontal rhombic antennas and four types of rhombic arrays; basic paper on rhombic arrays.
- _____, W. W. Jenvey, and R. D. Carmen. "Radio-Frequency Measurements on Rhombic Antennae," *AWA (Australia) Tech. Rev.*, vol. 7, no. 2 (1946): 131. Instrumentation and measuring techniques for determining current distributions on two-tier rhombic arrays.
- Colebrook, F. M. "Electric and Magnetic Fields for Linear Radiator Carrying a Progressive Wave," *J. IEE (London)*, vol. 89 (February 1940): 169.
- Dufour, J. "Diagrammes de réception d'antennes rhombiques dans un plan vertical: résultats expérimentaux," *Bull. Tech. PTT (Bern, Switzerland)*, vol. 31 (March 1953): 65. Measurements of vertical-plane patterns with aircraft, including effects of rough terrain.
- Harrison, C. W. Jr. "Radiation from Vee Antennas," *IRE Proc.*, vol. 31 (July 1943): 362.
- _____. "Radiation Field of Long Wires with Application to Vee Antennas," *J. App. Phys.*, vol. 14 (October 1943): 537.
- Iizuka, K. "Traveling-Wave V and Related Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-15 (March 1967).
- Laport, E. A. "Design Data for Horizontal Rhombic Antennas," *RCA Rev.*, vol. 13 (March 1952): 71.
- Wells, E. M. "Radiation Resistance of Horizontal and Vertical Aerials Carrying a Progressive Wave," *Marconi Rev.*, no. 83 (October–December 1946).

Chapter 11

Leaky-Wave
Antennas

Arthur A. Oliner

Polytechnic University

David R. Jackson

University of Houston

CONTENTS

11.1	INTRODUCTION	11-2
11.2	DESIGN PRINCIPLES FOR UNIFORM LEAKY-WAVE ANTENNAS	11-6
11.3	DESIGN PRINCIPLES FOR PERIODIC LEAKY-WAVE ANTENNAS	11-12
11.4	SPECIFIC STRUCTURES: OVERVIEW	11-15
11.5	SPECIFIC STRUCTURES BASED ON CLOSED WAVEGUIDES	11-16
11.6	SPECIFIC STRUCTURES BASED ON PERIODIC OPEN WAVEGUIDES	11-22
11.7	SPECIFIC STRUCTURES BASED ON UNIFORM OPEN WAVEGUIDES	11-29
11.8	ARRAYS THAT SCAN IN TWO DIMENSIONS	11-39
11.9	NARROW-BEAM ANTENNAS BASED ON A PARTIALLY REFLECTIVE SURFACE	11-44

11.1 INTRODUCTION

General Principles

A leaky-wave antenna is basically a waveguiding structure that possesses a mechanism that permits it to leak power all along its length. The earliest example of such an antenna is a rectangular waveguide with a continuous slit cut along its side,^{1,2} as shown in Figure 11-1. Since leakage occurs over the length of the slit in the waveguiding structure, the whole length constitutes the antenna's effective aperture unless the leakage rate is so great that the power has effectively leaked away before reaching the end of the slit. Because of the leakage, the leaky waveguide has a complex propagation wave number, with a phase constant β and a leakage constant α ; α is large or small depending on whether the leakage per unit length is large or small. A large α implies that the large leakage rate produces a short effective aperture, so that the radiated beam has a large beamwidth. Conversely, a low value of α results in a long effective aperture and a narrow beam, provided the physical aperture is sufficiently long.

When the antenna aperture is finite and fixed beforehand and the leakage rate α is small, the beamwidth is determined primarily by the fixed aperture, and the value of α influences the beamwidth only secondarily. What *is* affected strongly by the value of α under those conditions is the efficiency of radiation. We try to design a leaky-wave antenna so that its value of α allows about 90 percent of the power in the guide to be leaked away (radiated) by the time the wave reaches the end of the antenna aperture. The remaining power is absorbed by a matched load placed at the end of the waveguide. A typical leaky-wave antenna might be about 20 wavelengths long, so that the beamwidth of the radiation would be about 4° or so if the beam direction is about 45° from the leaky waveguide axis.

Because the phase constant β changes with frequency, the beam direction also changes with frequency, and the leaky-wave antenna can be *scanned* by varying the *frequency*. The precise ways in which changes in frequency affect the various properties of leaky-wave antennas are considered in detail later.

Since power is radiated continuously along the length, the aperture field of a leaky-wave antenna with strictly uniform geometry has an exponential decay (usually slow). Since the sidelobe behavior is poor for a leaky-wave antenna of typical length, the practice is then to vary the value of α slowly along the length in a specified way while maintaining β constant, so as to adjust the amplitude of the aperture distribution to yield the desired sidelobe performance. This tapering procedure is well known and is discussed later.

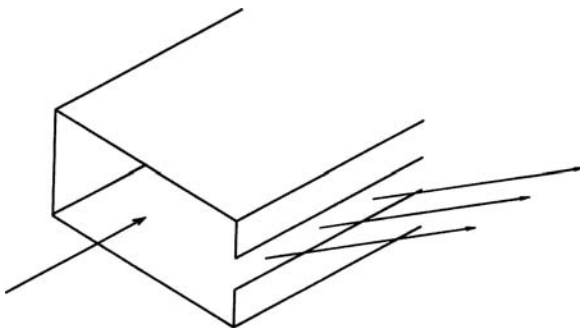


FIGURE 11-1 The earliest example of a leaky-wave antenna: a rectangular waveguide with a continuous slit cut along one of its sides

An individual leaky-wave antenna is clearly a line-source antenna. In three dimensions the beam is actually conical in shape, where the detailed beam shape depends on the cross-sectional dimensions of the leaky-wave antenna. Techniques are available for narrowing the beam in the cross-plane, such as the use of a horn or placing the line-source antenna in an array. Examples are given later.

As indicated earlier, the radiated beam of the leaky-wave antenna may be frequency scanned, producing a fan beam narrow in the scan plane. In Section 11.8 we describe how a linear array of such leaky-wave line sources may be used to produce a pencil beam and to permit independent scanning in *two dimensions*. The scanning in the cross-plane requires some mechanism other than a change in frequency, however, and phase scanning seems best. Examples of antennas that scan in two dimensions are presented near the end of this chapter in Section 11.8.

Leaky-wave antennas have been known and used for more than 40 years. Almost all the early antennas were based on closed waveguides that were made leaky by introducing a cut along the side of the waveguide (or something similar to that) to permit the power to leak away along the length of the waveguide. The newer millimeter-wave waveguides are actually *already open*, often in order to reduce the attenuation constants of such waveguides due to metal or dielectric losses. Examples are various kinds of dielectric waveguide, groove guide, NRD guide, microstrip line, etc.; of course, some of these guides are less lossy than others. The dominant modes on these open waveguides are generally purely bound, but a physical cut will not make them leak; instead, some new techniques are necessary, such as the introduction of asymmetry or some other modification of the geometry. Several examples are presented later.

The last remarks in this subsection relate to some confusion in the literature regarding the physics of leaky waves. The limited space available here will permit only a few words of explanation, but an examination of the wave numbers shows that the amplitude of a forward leaky wave *increases* transversely away from the guiding structure so that the wave violates the boundary condition at infinity in the transverse direction. Leaky waves have therefore been called “improper” or “nonspectral.” Although all this is true, these features of leaky waves do *not* complicate the design of leaky-wave antennas; the design principles, which are presented next, are actually quite simple and straightforward. Some simple considerations show that the leaky wave is actually defined only in a sector of space near the leaky-wave antenna and never reaches infinity in the transverse direction because the antenna itself is finite. The leaky-wave concept serves to provide the details of the aperture distribution and some of the main properties of the beam, but the radiation field itself is found in the usual fashion as the Fourier transform of the aperture field. There are some sophisticated mathematical aspects regarding the improper (nonspectral) nature of leaky waves,³⁻⁵ but some of the considerations referred to earlier in this paragraph can be explained in simple geometric terms.²

Two Types of Leaky-Wave Antennas: Uniform and Periodic

There are two different basic types of leaky-wave antennas, depending on whether the geometry of the guiding structure is uniform or is periodically modulated along its length. These two types are actually similar in principle to each other, but their performance properties differ in several ways, and they face somewhat different problems in their design. The two types are therefore treated separately in the discussions that follow.

The *first* type, the *uniform* leaky-wave antenna, is uniform along the length of the guiding structure, as opposed to possessing some periodic modulation. (As mentioned earlier, we recognize that the uniform leaky-wave antenna has a small taper along its length in order to improve and control the sidelobe level.)

All leaky-wave line sources of the uniform type radiate into the forward quadrant and can scan in principle from broadside to end fire, with the beam nearer to end fire at the higher frequencies. In practice, however, you cannot get too close to end fire or to broadside, but how near those limits can be approached depends on the specific structure. For example, suppose the cross section of the guiding structure contains dielectric material in part and air in part, and has a slow-wave range ($\beta > k_0$) and a fast-wave range ($\beta < k_0$), where k_0 is the free-space wave number in air. Then the transition between the two ranges is usually a rapid one, occurring at end fire (when $\beta = k_0$), and the beam can be scanned very close to end fire. An additional virtue of this structure is that a wide scan-angle range can usually be covered with only a relatively small frequency range. An example is given later. When the structure is filled with air only, on the other hand, you must stay about 10 or 15° away from both broadside and end fire, and the frequency sensitivity is more sluggish, particularly near end fire. An important virtue possessed by such single-medium leaky-wave antennas, however, is that the beamwidth remains exactly constant as the beam is scanned with frequency.

In the *second* type of leaky-wave antenna, the *periodic* type, some periodic modulation of the guiding structure is introduced, and it is this periodicity that produces the leakage. The periodic modulation itself is uniform along the structure's length, again except for the small taper of the periodic properties along the length to control the sidelobes. Again, a complex propagation wave number results, with β and α ; large or small values of α are related to the beamwidth and the radiation efficiency in the same manner as that found for uniform leaky-wave antennas. A typical example of a periodic leaky-wave antenna is a dielectric rectangular rod on which a periodic array of metal strips is placed, as seen in Figure 11-2.

An important difference between uniform and periodic leaky-wave antennas is that the dominant mode on the former is a fast wave that therefore radiates whenever the structure is open. On the other hand, the dominant mode on a periodic leaky-wave antenna is a slow wave that does not radiate even though the structure is open. Introduction of the periodic array produces an infinity of space harmonics, some of which may be fast while the rest

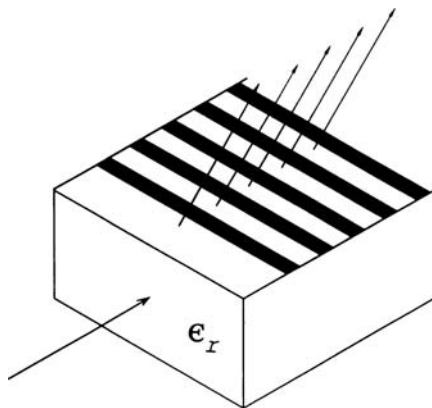


FIGURE 11-2 A typical and important example of a periodic leaky-wave antenna: a rectangular dielectric rod (which may or may not be situated on a metal plane) on which is placed a periodic array of metal strips. This antenna can radiate into either the forward or backward quadrants.

are slow; the fast space harmonics would radiate. Since we desire an antenna that radiates only a single beam, the structure is designed so that only the first space harmonic ($n = -1$) is fast. The relevant design considerations are presented in Section 11.3. It is also shown there that the scan range for this class of antennas is from backward end-fire through broadside into part of the forward quadrant, except for a narrow region around broadside, where an “open stop band” occurs.

In general terms, therefore, we see that the scan range is completely different from that for the uniform leaky-wave antennas. There the beam scans in the forward quadrant only; also, it cannot approach broadside too closely, and sometimes it cannot be used too near to end fire. For the periodic leaky-wave antenna, we can scan over almost all the backward quadrant and into some of the forward quadrant as well.

Relation to Surface-Wave Antennas and Slot Arrays

Surface-wave antennas, leaky-wave antennas, and slot arrays are all members of the family of traveling-wave antennas, yet they are treated in this *Handbook* in separate chapters (Chapters 10, 11, and 9, respectively). They are similar to each other in some evident respects (e.g., the basic structure in each case is a waveguide of some sort), but they all differ from each other in important ways that lead to different design procedures and to different performance expectations.

Surface-wave antennas are purely end-fire antennas, whereas leaky-wave antennas and slot arrays do not radiate well in the end-fire direction and, in fact, are designed either to radiate in some other direction or to scan over a range of angles. The basic guiding structure for surface-wave antennas is an open waveguide (such as a dielectric rod) whose dominant mode (the surface wave) is purely bound, so that the surface wave will radiate only at discontinuities, such as the very end of the waveguide. It does not radiate along the length of the guide because the surface wave is a slow wave (except for some small leakage into the almost-end fire direction if the surface-wave antenna is tapered), whereas a uniform leaky-wave antenna, which supports a fast wave, leaks power all along the length of the waveguide.

Periodic leaky-wave antennas, however, are often based on surface waveguides. There the basic waveguide can be an open structure whose dominant mode is a surface wave that is a slow wave, and the radiation is produced by placing a periodic array of discontinuities on the guide in such a way that the first space harmonic becomes fast. The resulting leaky wave may therefore be viewed as arising from a *surface-wave-excited array*. There is thus a strong kinship between the basic structures employed for surface-wave antennas and those for periodic leaky-wave (or surface-wave-excited) arrays, and, for this reason, periodic leaky-wave antennas were, in the first two editions of this *Handbook*, included within the same chapter as surface-wave antennas. On the other hand, since the performance properties of periodic leaky-wave antennas, and the design procedures to achieve those properties, are very different from those for surface-wave antennas but very similar to those for uniform leaky-wave antennas, the leaky-wave antennas of both types are currently incorporated into the present chapter.

The differences between *periodic* leaky-wave antennas and *slot arrays* are more subtle but still very significant. A visually evident difference is that most slot arrays are fed from air-filled rectangular waveguides whose dominant mode is fast; to suppress the radiation from this fast dominant ($n = 0$) space harmonic while retaining that from the $n = -1$ space harmonic, it is necessary to place successive slots on alternate sides of the guide centerline or to alternately tilt the slots to produce phase reversals. Periodic leaky-wave antennas do not need such phase reversals because the (slow) $n = 0$ space harmonic does not radiate. However, this distinction is *not* fundamental, as may be noted when the slot-array rectangular waveguide is dielectric-filled and the alternation of slots is no longer needed. The structures then resemble each other in principle.

The basic distinction between periodic leaky-wave antennas and slot arrays lies in the nature of the individual discontinuities, and, therefore, in the basic design approaches. The intention in the leaky-wave antennas is to produce a slow leakage per unit length; thus each discontinuity element in the periodic array of elements produces a small loading on the basic waveguide mode. The individual elements are intentionally made nonresonant. As a result, the design procedure views the leaky-wave antenna as an equivalent “smooth” structure with a complex propagation wave number, where the array of discontinuity elements is considered as a whole in the analysis.

In contrast, the slots in a slot array are considered *individually*, and then mutual coupling effects are taken into account when the array itself is formed. Furthermore, the individual slots are usually *resonant*. Variations in the loading on the basic waveguide mode are achieved by rotating the slots or by moving them closer to the guide centerline. The design approach thus becomes quite different from that for leaky-wave antennas, thereby warranting inclusion in a separate chapter.

The resonant loading in slot arrays, as opposed to the nonresonant loading in periodic leaky-wave antennas, also influences performance. For example, slot arrays are more frequency-dependent, which can be good if you wish to scan more rapidly with frequency, but the trade-off is that the scan range is narrowed. Most slot arrays are not intended for a large scan range, however. On the other hand, if slot arrays are built with nonresonant slots, their behavior would be very similar to that of periodic leaky-wave antennas. The question then would be whether the customary slot-array design or the leaky-wave design would be more suitable. One last observation relates to the fact that the treatment in Chapter 9 is restricted to slot arrays fed by rectangular waveguide. The design approach described for such arrays could be applied equally well to arrays of other resonant elements, e.g., the two-wire line with proximity-coupled dipoles.⁶

11.2 DESIGN PRINCIPLES FOR UNIFORM LEAKY-WAVE ANTENNAS

The physical structure of a leaky-wave antenna consists of a leaky waveguide with a length L along which the leakage occurs. The propagation characteristics of the leaky mode in the longitudinal (z) direction are given by phase constant β and leakage constant α , where α is a measure of the power leaked (and therefore radiated) per unit length. The length L then forms the aperture of the line-source antenna, and the amplitude and phase of the traveling wave along the aperture are determined by the values of β and α as a function of z . When the leaky waveguide is completely uniform along its length, β and α do not change with z , and the aperture distribution has an exponential amplitude variation and a constant phase. For a leaky-wave antenna of typical length, such an aperture distribution results in a high sidelobe level, so that the design of a practical leaky-wave antenna will include a variation of α with z in order to control the sidelobes in some specified fashion. More is said about this later in this section.

The values of β and α will depend on the precise cross-sectional geometry of the leaky waveguide, and the determination of β and α , whether theoretically or experimentally, is in most cases the most difficult part of the design. Their knowledge, however, is essential to any systematic design procedure.

Once β and α are known as a function of frequency and cross-sectional geometry, the principal behavioral features of a leaky-wave antenna follow very quickly. Such features include the beam direction, the beamwidth, the radiation efficiency, the variation of the scan angle with frequency, and the taper in α required to control the sidelobes.

Beam Direction, Beamwidth, and Radiation Efficiency

These major behavioral features follow directly once the values of β and α are known, and they are given to a good approximation by a set of very simple relations. We first consider the beam direction and the beamwidth:

$$\sin \theta_m \approx \frac{\beta}{k_0} \quad (11-1)$$

$$\Delta\theta \approx \frac{1}{(L/\lambda_0)\cos\theta_m} \quad (11-2)$$

Here θ_m is the angle of the maximum of the beam, measured from the broadside direction (perpendicular to the leaky-wave antenna axis), L is the length of the leaky-wave antenna, $\Delta\theta$ is the beamwidth, and k_0 is the free-space wave number ($= 2\pi/\lambda_0$). Both θ_m and $\Delta\theta$ are in radians in Eqs. 11-1 and 11-2. The beamwidth $\Delta\theta$ is determined primarily by the antenna length L , but it is also influenced by the aperture field amplitude distribution. It is narrowest for a constant aperture field and wider for sharply peaked distributions. Equation 11-2 is a middle-of-the-range result. For a constant aperture distribution, the unity factor in the numerator should be replaced by 0.88; for a leaky-wave structure that is maintained uniform along its length, consistent with 90 percent radiation, the factor should be 0.91; for a tapered distribution that is sharply peaked, the factor could be 1.25 or more (see Table 1-1 for examples).

The antenna length L is usually selected, for a given value of α , so that 90 percent (or at most 95 percent) of the power is radiated, with the remaining 10 percent or so absorbed by a matched load. Attempting to radiate more than 90 percent or so creates two problems: First, the antenna must be made longer, and second, the variation in $\alpha(z)$ required to control the sidelobes becomes extreme. For 90 percent of the power radiated, we find

$$\frac{L}{\lambda_0} \approx \frac{0.18}{\alpha/k_0} \quad (11-3)$$

This simple but useful relation follows from writing

$$\frac{P(L)}{P(0)} = \exp(-2\alpha L) = \exp[-4\pi(\alpha/k_0)(L/\lambda_0)] \quad (11-4)$$

where $P(L)$ is the power remaining in the leaky mode at $z = L$ and $P(0)$ is the power input at $z = 0$.

If both L and α are specified independently, the percentage of power radiated can deviate significantly from the desired 90 percent. In fact, α is a function of frequency, so that the radiation efficiency will change somewhat as the beam is frequency scanned. The 90 percent figure is usually applied to the middle of the scan range. Using Eq. 11-4, however, we can easily obtain an expression for the percentage of power radiated:

$$\begin{aligned} \text{Percentage of power radiated} &= 100[1 - P(L)/P(0)] \\ &= 100\{1 - \exp[-4\pi(\alpha/k_0)(L/\lambda_0)]\} \end{aligned} \quad (11-5)$$

Equation 11-5 assumes an exponentially decaying aperture distribution. If the aperture distribution has been changed in order to control sidelobes, as is customary, Eq. 11-5 is still useful as a good approximation.

Scan-Angle Behavior

There are two different types of uniform leaky-wave antennas that are similar in principle but that differ somewhat in their scan-angle behavior. The guiding structures for these two types differ in that they are *air-filled* for one type and *partially dielectric-filled* for the other.

Typical air-filled guiding structures would include open rectangular waveguide and groove guide, for which the dominant modes are fast relative to the free-space velocity. Guiding structures that are partially dielectric-filled include nonradiative dielectric (NRD) guide and open dielectric-loaded rectangular waveguide. Depending on the frequency and the geometry, the dominant modes on these guiding structures can be fast or slow, but when they are used as leaky-wave antennas, it is necessary to operate them in the fast-wave range ($\beta/k_0 < 1$), of course.

There are advantages and disadvantages in performance when the guiding structures are air-filled or when they are partially loaded with dielectric. With respect to the variation of beamwidth with scan angle, the air-filled structures are superior. Because the transverse wave number is then a constant, independent of frequency, the beamwidth of the radiation remains exactly constant as the beam is scanned by varying the frequency. With partial dielectric loading, on the other hand, the beamwidth changes with scan angle. With respect to frequency sensitivity, i.e., how quickly the beam angle scans as the frequency is varied, the partly dielectric-loaded structure can scan over a larger range of angles for the same frequency change and is therefore preferred.

The reason for these differences in behavior between the air-filled and partly dielectric-loaded cases is shown in Figure 11-3. Let us first recall a few features. We begin with Eq. 11-1, where θ_m is the angle of the maximum of the beam, measured from broadside.

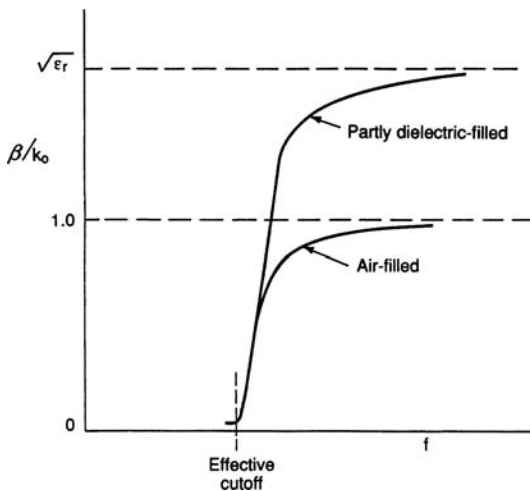


FIGURE 11-3 The variations of the normalized wave number β/k_0 with frequency for uniform leaky-wave antennas that are air-filled or are partly filled with dielectric material. These variations explain their different scan-angle behaviors.

Then we note that the line $\beta/k_0 = 1$ corresponds to end fire, and $\beta/k_0 = 0$ corresponds to broadside. The β/k_0 variation near to cutoff (near to broadside) is seen to be much the same whatever the filling factor of the guiding structure. The big difference occurs near end fire. The variation with frequency for the air-filled case is seen to be quite slow near the $\beta/k_0 = 1$ line because the curve asymptotically approaches that line as the frequency becomes large. For the partly dielectric-filled case, on the other hand, the curve goes quite rapidly to the $\beta/k_0 = 1$ line and above it. As a result, the variation of scan angle with frequency is more rapid overall for the partly dielectric-filled case, and we can approach end fire rather closely in the scan-angle range. In contrast, we cannot approach end fire closely when the guiding structure is air-filled.

We indicated earlier that for air-filled guiding structures the beamwidth $\Delta\theta$ remains constant as the beam is scanned by varying the frequency. This statement is easily proved once we recall that for such air-filled structures the transverse wave number k_t is a constant (k_c) independent of frequency. Using Eq. 11-1, we find

$$\cos^2 \theta_m = 1 - \sin^2 \theta_m = 1 - (\beta/k_0)^2$$

Then, since

$$k_0^2 = \beta^2 + k_c^2$$

for air-filled guiding structures, we may write

$$1 - (\beta/k_0)^2 = 1 - [1 - (k_c/k_0)^2] = (k_c/k_0)^2$$

so that

$$\cos \theta_m = k_c/k_0 \quad (11-6)$$

Substituting Eq. 11-6 into Eq. 11-2 yields

$$\Delta\theta \approx \frac{2\pi}{k_c L} = \frac{\lambda_c}{L} \quad \text{in radians} \quad (11-7)$$

which is independent of frequency, provided the antenna is designed according to Eq. 11-3. When the guiding structure is partly filled with dielectric, the transverse wave number k_t is a function of frequency, so that $\Delta\theta$ changes as the beam is frequency scanned.

Radiation Pattern

As usual, the radiation pattern can be found by taking the Fourier transform of the aperture distribution. When the geometry of the leaky-wave antenna is maintained constant along the antenna length, the aperture field distribution consists of a traveling wave with a constant β and α , meaning that the amplitude distribution is exponentially decaying. If the antenna length is infinite, the radiation (power) pattern $R(\theta)$ is given by

$$R(\theta) \sim \frac{\cos^2 \theta}{(\alpha/k_0)^2 + (\beta/k_0 - \sin \theta)^2} \quad (11-8)$$

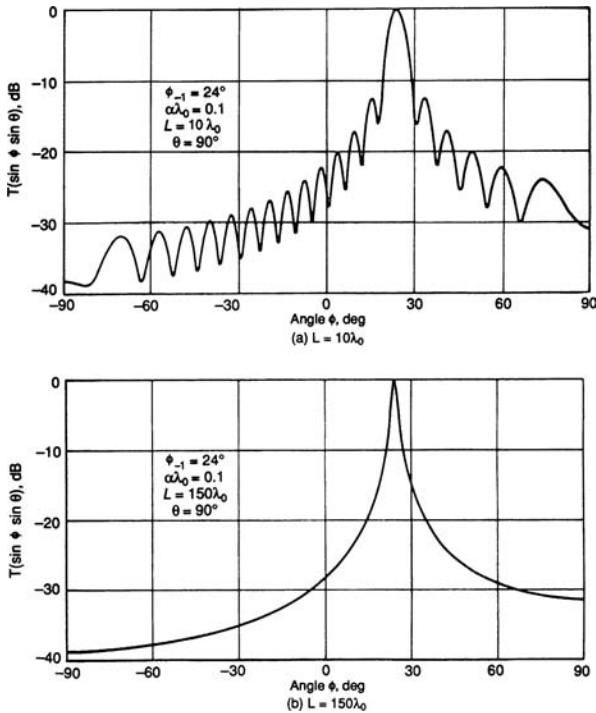


FIGURE 11-4 Radiation patterns of dielectric grating leaky-wave antennas, showing the changes in the sidelobe behavior with antenna length. These antennas are not tapered to control the sidelobes. (a) Antenna length $L = 10\lambda_0$ (sidelobes clearly present). (b) Antenna length $L = 150\lambda_0$ (no sidelobes) (after Schwering and Peng⁷ © IEEE 1983).

which does not exhibit any sidelobes. If the antenna length is finite, the expression for $R(\theta)$ becomes more involved, and the pattern possesses sidelobes that modify the basic shape for infinite length.

The preceding comments are illustrated well in a paper on dielectric-grating leaky-wave antennas by Schwering and Peng.⁷ They present several examples of such patterns, two of which are shown in Figure 11-4a and b. The length of the antenna in Figure 11-4a, which clearly exhibits sidelobes, is $10\lambda_0$. As the antenna length increases, the amplitude of the sidelobe variations decreases. For the radiation pattern in Figure 11-4b, which shows a smooth pattern only, the antenna length is $150\lambda_0$, which is evidently effectively infinite. Although the leaky-wave antenna for which these calculations were made is periodic rather than uniform, the basic features are identical. Schwering and Peng⁷ contains an extended discussion of radiation-pattern considerations, including equations and other figures.

The radiation pattern in Figure 11-4a is seen to possess first sidelobes that are only about 13 dB down, which is generally undesirable. To greatly reduce the sidelobe level and to control the pattern in other ways, it is customary to appropriately taper the amplitude of the aperture distribution, as is discussed next.

Control of Aperture Distribution to Reduce Sidelobes

The procedure to design the leaky-wave antenna so that it produces a final desired radiation pattern is straightforward, though somewhat complicated, involving the following steps. First, the final desired radiation pattern is specified, and then the corresponding aperture amplitude distribution is determined by standard antenna techniques. Then, by using the expression derived next, the values of α/k_0 are computed as a function of position along the antenna length in accordance with the aperture amplitude distribution that was just determined. At the same time β/k_0 must be maintained constant along the length so that the radiation from all parts of the aperture point in the same direction. Finally, from the theory that relates α and β to the geometry of the structure, we compute the tapered geometry as a function of position along the antenna length.

When we change the local cross-sectional geometry of the guiding structure to modify the value of α at some point z , however, it is likely that the value of β at that point is also modified slightly. However, since β must not be changed, the geometry must be further altered to restore the value of β , thereby changing α somewhat as well. In practice, this difficulty requires a two-step process for most leaky-wave antennas, which is not bad. Because of this added complexity, we seek leaky-wave structures for which we can vary geometric parameters that change β and α essentially independently.

The first design step mentioned earlier, i.e., determining the required aperture amplitude distribution for the selected desired radiation pattern, is a standard antenna procedure not specifically related to leaky-wave antennas. The second step, calculating the value of $\alpha(z)$ corresponding to the aperture amplitude distribution found from the first step, is directly pertinent to leaky-wave antennas, and we therefore present now a derivation of the expression needed for the second step.

The power distribution along the antenna can be expressed in the form

$$P(z) = P(0) \exp \left[-2 \int_0^z \alpha(\zeta) d\zeta \right] \quad (11-9)$$

where $P(0)$ is the power at the input point, $z = 0$, and ζ is the integration variable. Upon differentiation of Eq. 11-9, we obtain

$$-\frac{dP(z)}{dz} = 2\alpha(z)P(z) \quad (11-10)$$

Suppose now that the desired aperture distribution (which would achieve the specified radiation pattern) is $A(z) \exp(-j\beta z)$. We may then write

$$-\frac{dP(z)}{dz} = c |A(z)|^2 \quad (11-11)$$

where c is a constant of proportionality. Comparison of Eqs. 11-10 and 11-11 yields

$$2\alpha(z) = \frac{c |A(z)|^2}{P(z)} \quad (11-12)$$

Upon integration of Eq. 11-11, we obtain, corresponding to two sets of limits of integration, the following:

$$c \int_0^L |A(\zeta)|^2 d\zeta = P(0) - P(L) \quad (11-13)$$

$$c \int_0^z |A(\zeta)|^2 d\zeta = P(0) - P(z) \quad (11-14)$$

We next use Eq. 11-14 to substitute for $P(z)$ in Eq. 11-12, and then we employ Eq. 11-13 to eliminate the proportionality constant c . In a straightforward fashion, we then obtain the desired result:

$$2\alpha(z) = \frac{|A(z)|^2}{\frac{P(0)}{P(0) - P(L)} \int_0^L |A(\zeta)|^2 d\zeta - \int_0^z |A(\zeta)|^2 d\zeta} \quad (11-15)$$

The units of $\alpha(z)$ in Eq. 11-15 are nepers per unit length. To obtain $\alpha(z)$ in decibels per unit length, we multiply by 8.68. If $P(L)$, the power remaining at the end of the aperture, is allowed to approach zero, we note from Eq. 11-15 that $\alpha(z)$ then becomes very large for points near to the end of the aperture, i.e., for z approaching L . This is the main reason why it is common for $P(L)/P(0)$ to be equal to 0.1 or so, but not much smaller, with the remaining power being absorbed in a matched load to avoid the presence of any backlobe.

11.3 DESIGN PRINCIPLES FOR PERIODIC LEAKY-WAVE ANTENNAS

As discussed in Section 11.1, periodic leaky-wave antennas differ from uniform ones in that the waveguiding structure is modulated periodically along its length instead of being completely uniform (again, except for the small taper for both types to control the side-lobes). The dominant mode on uniform antennas is fast relative to free-space velocity, whereas the one on periodic antennas is slow, so that the dominant mode itself does not radiate and it needs the periodic modulation to produce the radiation. Since the physical processes that produce the radiation are different, these two antenna types have different scan ranges. On the other hand, most of the design principles for the uniform leaky-wave antennas discussed in Section 11.2 also apply to the periodic ones. The treatment given next indicates in what ways changes in design are necessary. First, however, we summarize how the periodicity produces the leakage and, in that context, why the scan ranges are different for the two types.

Effect of Periodicity on Scan Behavior

To explain the source of the leakage and to understand the scan behavior as a function of frequency, we invoke the concept of *space harmonics*. Suppose we first take a uniform dielectric waveguide, and then we place an array of metal strips periodically along its length (as in Figure 11-2). Before the metal strips are added, we choose the guide dimensions and frequency so that only the dominant mode is above cutoff; furthermore, $\beta > k_0$ for this mode, so it is purely bound. When the periodic array of strips is added, the periodicity

introduces an infinity of space harmonics, each characterized by phase constant β_n and related to each other by

$$\beta_n d = \beta_0 d + 2n\pi \quad (11-16)$$

where d is the period and β_0 , the fundamental space harmonic, is simply the original β of the dominant mode of the uniform dielectric waveguide, but perturbed somewhat in value because of the addition of the strips. As seen from Eq. 11-16, β_n can take on a large variety of values, so that these space harmonics can be forward or backward in nature, and be slow or fast. Since the structure is open, a space harmonic that is fast will radiate. To say it in another way, since the space harmonics are all tied together, and all of them together comprise the dominant mode of the loaded structure, the whole mode becomes leaky if one or more of the space harmonics becomes fast.

We recall that for a space harmonic to be fast, we need $\beta_n/k_0 < 1$; we also know that $\beta_0/k_0 > 1$. If we rewrite Eq. 11-16 in the form

$$\frac{\beta_n}{k_0} = \frac{\beta_0}{k_0} + \frac{2n\pi}{k_0 d} = \frac{\beta_0}{k_0} + \frac{n\lambda_0}{d} \quad (11-17)$$

we see that $|\beta_n/k_0|$ can readily be less than unity if n is negative and λ_0/d is suitably chosen. For a practical antenna, we want only a single radiated beam, so we choose $n = -1$.

If we follow this line of thinking further (the details will not be given here), we will find that when the frequency is low, all the space harmonics are slow and there are no radiated beams. When the frequency reaches the critical value for which the $n = -1$ space harmonic first becomes fast, the radiating beam just emerges from backward end-fire. As the frequency is increased, the beam swings up from backward end-fire but is still radiating into the backward quadrant. Further increases in the frequency will swing the beam toward broadside, then through broadside, and finally into the forward quadrant. The amount by which the forward quadrant is penetrated depends on other properties of the antenna, in this case primarily on the dielectric constant. On the other hand, the antenna is useful only if a single, controllable beam is radiated, and the range in the forward quadrant is usually limited by the emergence of the $n = -2$ beam from backward end-fire, or by the next waveguide mode coming above cutoff.

A special problem is present for the periodic structures near broadside. A narrow region around broadside corresponds to an “open stopband” region, where the value of α becomes large and then zero (for a structure of infinite length). In a practical antenna, this means that within this narrow angular region the amount of radiation drops substantially, and a large VSWR is encountered (power is reflected back to the source rather than being radiated). This effect is well known, and it also occurs when slot arrays are scanned through broadside. There are techniques, not widely used, that permit these arrays to scan through broadside, however. One of them^{8,9} uses pairs of strips rather than single strips, where the spacing between the elements of each pair is $\lambda_{g0}/4$ (where λ_{g0} is the guide wavelength) at the broadside frequency, so that the wave reflected at broadside by the first element of each pair will be canceled, or nearly so, by the wave reflected by the second element.

Beam Direction, Beamwidth, and Radiation Efficiency

The discussion in Section 11.2 under this heading shows that these major behavioral features are given to a good approximation by a set of very simple relations. All the considerations

presented there for uniform leaky-wave antennas apply as well to periodic leaky-wave antennas provided that we make one simple change, which takes into account the main difference between the two antenna types. The main difference relates to the fact that radiation from *periodic* leaky-wave antennas is due to the $n = -1$ space harmonic. In Eq. 11-1, therefore, β must be replaced by β_{-1} , to yield

$$\sin \theta_m \approx \frac{\beta_{-1}}{k_0} \quad (11-18)$$

where

$$\beta_{-1} = \beta_0 - 2\pi/d \quad (11-19)$$

consistent with Eq. 11-16. When we substitute Eq. 11-19 into Eq. 11-18, we obtain

$$\sin \theta_m \approx \frac{\beta_0}{k_0} - \frac{2\pi}{k_0 d} = \frac{\lambda_0}{\lambda_{g0}} - \frac{\lambda_0}{d} \quad (11-20)$$

Thus, depending on how λ_0/d , where d is the period, compares with λ_0/λ_{g0} (or β_0/k_0), the beam can point in the backward quadrant or in the forward quadrant, in accordance with the discussion in the preceding subsection.

Equations 11-2 through 11-6, and the discussions relating to them, also apply to periodic leaky-wave antennas when the distinction relating to β_{-1} is kept in mind.

The considerations in Section 11.2 relating to the radiation pattern, as well as to the steps required to control the aperture distribution, are also valid here.

Feed Considerations

When the antenna aperture is tapered appropriately in accordance with the design steps outlined in Section 11.2 in the subsection “Control of Aperture Distribution to Reduce Sidelobes,” very beautiful radiation patterns, with very low sidelobes, can be achieved on paper. These fine results can also be obtained in practice when proper attention is also paid to the way the leaky-wave antennas are fed.

For uniform leaky-wave antennas that are formed by opening up initially closed waveguides, concerns relating to the feed are usually negligible or nonexistent. The reason for this lies in the taper process for controlling the sidelobes. The taper is such that the aperture radiates very little at its ends, and therefore the discontinuity between the closed feed waveguide and the antenna aperture when it first begins is extremely small in most cases. There is therefore negligible spurious radiation from the feed junction. There is also no appreciable impedance mismatch at that feed junction, of course.

When the feeding structure is an open waveguide, we must examine the situation more carefully. This is particularly true for those periodic leaky-wave antennas that are basically surface-wave-excited. The problem then lies with the way the basic surface wave is produced, rather than with the transition to the periodic modulation, which again begins very slowly due to the taper. Surface waves are often formed by a tapered transition from a closed waveguide, with resultant spurious radiation associated with the transition. Such problems are well known with respect to surface-wave antennas. When such transitions form part of the overall feed system, their contributions to the radiation pattern may be significant and may spoil the initial careful paper design.

In many cases, spurious feed radiation is not a problem with leaky-wave antennas, and this is one of their important virtues, but the feed mechanism must be looked at

carefully in the design to make sure that it does not introduce its own contribution to the radiation pattern.

11.4 SPECIFIC STRUCTURES: OVERVIEW

Leaky-wave antennas are formed by perturbing an initially bound mode on a waveguiding structure in a way that produces leakage all along the length of the guiding structure. As might be expected, all the very early leaky-wave antennas were based on *closed waveguides*, and the leakage was achieved by physically cutting into a wall of the waveguide, in the form of a longitudinal slit or a series of closely spaced holes. The first known leaky-wave antenna, shown in Figure 11-1 and invented by W. W. Hansen¹ in 1940, was in fact a long slot in a rectangular waveguide.

The next section presents several of the more important leaky-wave antennas based on *closed* waveguides. These structures are usually very simple in cross section, so that it was possible to obtain accurate expressions for the complex wave number ($\beta - j\alpha$) in terms of the frequency and the geometry of the cross section. Since very little spurious radiation occurs because of the feed junction, the theoretically derived radiation patterns, based on designs that yielded low sidelobes, agreed extremely well with the measured radiation patterns. A dramatic example of excellent agreement is given in Section 11.5. Many of the early theoretical expressions and the design based on them are still of practical value today and are indeed so good that the initial designs were often the final ones, without the need for any empirical adjustments.

The data relating the wave-number behavior to the geometric parameters are usually plotted as β/k_0 or $\lambda_0\lambda_g$, where λ_g is the guide wavelength in the propagation direction (z), and as α/k_0 or $\alpha\lambda_0$, where α is the leakage constant. In previous editions of this *Handbook*, plots were presented in the form of λ/λ_z and $\alpha_z\lambda$, where α_z , λ , and λ_z were written instead of α , λ_0 , and λ_g . Where the discussion and graphical plots in the present chapter employ material taken from the previous edition, the previous notation is carried over. It is essential that such information be available for any specific antenna structure because the correct taper for low sidelobes cannot be designed without it. A last general remark to be made in connection with leaky-wave structures based on closed waveguides is that some of the structures employ a series of round holes or small (nonresonant) slots. These holes or slots, however, are closely spaced, so that the structures should be viewed as quasi-uniform rather than periodic, even though these holes are periodically spaced. The radiation produced by them radiates the $n = 0$ space harmonic and not the $n = -1$ space harmonic.

The next stage in the development of specific leaky-wave antennas involved those based on *open* waveguides. Some of these antennas are uniform structures that employ open waveguides on which the dominant mode is initially purely bound, and others are periodic structures that are excited by surface waves and radiate via the $n = -1$ space harmonic. These two categories are considered separately in Sections 11.7 and 11.6, respectively.

The best-known examples of surface-wave-excited *periodic* leaky-wave antennas are dielectric rectangular rods (or slabs), with or without ground planes, that have on their tops or on one of their sides a periodic array of grooves or a periodic array of metal strips. Another large group is based on microstrip line. These antennas have been studied rather extensively, both experimentally and, more recently, theoretically. Together with other, similar structures, they are discussed in Section 11.6. An important problem for this class of structures is the incorporation of a feed mechanism that does not contribute spurious radiation.

Uniform leaky-wave antennas based on *open* waveguides offer a special challenge. Since the guide is already open, it cannot be cut to induce radiation, and other approaches

are needed. The most common one is the appropriate introduction of asymmetry, but other mechanisms, such as the use of a leaky higher-order mode or some modification in the geometry, have also been found useful. The first (and only early) example of such a leaky-wave antenna based on an open waveguide was invented by W. Rotman in the late 1950s. This pioneering study^{10,11} involved several versions of a form of trough waveguide whose dominant mode is purely bound but was made leaky by introducing asymmetry. The design procedure for this antenna was successful and practical, but the approach was not pursued further until about 20 years later, in the context of a new need that emerged in connection with millimeter waves.

As interest in millimeter waves revived during the 1970s, it was recognized that new forms of leaky-wave antennas were needed because of the smaller wavelengths involved and because the usual waveguides had higher loss at those higher frequencies. Since these smaller wavelengths caused fabrication problems due to small dimensions, simpler structures were sought; in fact, the latest structures are designed to permit the complicated portion of the structure, including the taper for sidelobe control, to be deposited photolithographically, in printed-circuit form, by using a mask. Because of the loss considerations, the newer antennas are often based on new lower-loss waveguides designed for millimeter-wave applications.

These include nonradiative dielectric (NRD) guide, groove guide, and rectangular dielectric rods, sometimes used in conjunction with microstrip in novel ways. Unfortunately, these last-mentioned structures were difficult to analyze theoretically, so their designs were empirical only; as a result, the experimental radiation patterns showed poor sidelobe performance, leading some people to conclude unfairly that poor sidelobe behavior was a necessary consequence of this class of antennas. However, many other leaky-wave antennas in this class were analyzed accurately, primarily by Oliner and his colleagues, and their results agreed very well with measurements. Some of the more promising of these newer millimeter-wave antennas are described in Section 11.7.

A more recent development relating to leaky-wave antennas is their incorporation into arrays that permit scanning in two dimensions. The arrays are essentially a linear phased array of leaky-wave line-source antennas, where the scanning in elevation is obtained in leaky-wave fashion by varying the frequency, and the scanning in azimuth is achieved by varying the phase difference between the successive parallel leaky-wave line sources. The architecture underlying this approach is described in Section 11.8, and several examples are given of specific antenna structures in this category. A partial motivation for this approach is to achieve a lower-cost substitute for phased-array antennas in some applications.

11.5 SPECIFIC STRUCTURES BASED ON CLOSED WAVEGUIDES

The earliest example of a leaky-wave antenna was the one for which W. W. Hansen was granted a patent.¹ He had proposed during the late 1930s that an antenna could be created by cutting a rectangular waveguide longitudinally, as shown in Figure 11-1, thereby producing a long slit in the side of the initially closed guide, out of which power could leak away. His concept was not pursued at that time because of the success of slot arrays, but the simplicity of the structure remained attractive, and it was reexamined about a decade later. The 1950s, in fact, represented a very active period during which many leaky-wave antennas based on closed waveguides were proposed, analyzed, measured, and utilized. Several excellent references summarize in detail the state of the art in this class of antennas as of the middle 1960s, including a comprehensive book¹² by C. H. Walter; a chapter by F. J. Zucker¹³ in the First Edition of this *Handbook*; a chapter by T. Tamir¹⁴ in Part II of

the book *Antenna Theory*, edited by R. E. Collin and F. J. Zucker; and a summary by A. A. Oliner and R. G. Malech¹⁵ in Volume II of the book *Microwave Scanning Antennas*, edited by R. C. Hansen.

All these specific structures based on closed waveguides are “uniform” leaky-wave antennas, so the principles for their design are those discussed earlier in Section 11.2. The remaining information required to complete the design involves the expressions for β/k_0 ($=\lambda/\lambda_z$) and α/k_0 (or $\alpha_z\lambda$) as a function of the frequency and geometric parameters of the specific structure. Since the period before the middle 1960s predates the computer era, theoretical expressions had to be simple to be considered practical. Fortunately, the structures themselves were simple, leading automatically to relatively simple expressions that were accurate, but, in addition, many of the expressions were further simplified by the use of perturbation relations.

In some cases, these values of α and β were measured rather than calculated, and then employed in the design. Two experimental methods are outlined in Walter’s book.¹⁶

Long Slits in Rectangular Waveguide

Radiation from long slits in rectangular waveguide can be accomplished in several ways, where the leakage rate can be adjusted by changing the slit width, and the polarization of the radiated beam can be selected by changing the waveguide mode. Some examples are illustrated in Figure 11-5.

The best-known example in this well-known category is the narrow slit in the side wall of rectangular waveguide shown in Figure 11-1, or the same structure with a ground plane seen in Figure 11-5a. The antenna shown in Figure 11-5a also differs from the structure in Figure 11-1 in that it is rotated by 90° and the slit is shown tapered (in exaggerated fashion) to remind us that in the design the slit width is varied to control the sidelobes in the radiation pattern. (The rectangular waveguide dimensions h and w correspond, of course, to the usual a and b , respectively.)

For this antenna, with the ground plane present but with the slit uniform, simple theoretical expressions are available for the relations between λ/λ_z and $\alpha_z\lambda$ and the frequency and the geometry. These expressions were derived by Goldstone and Oliner¹⁷ using a transverse resonance approach together with a simple perturbation procedure. The results agreed very well with measurements. The expressions (with the notation differing somewhat from that in Goldstone and Oliner¹⁷ to be consistent with Figure 11-5a) are

$$\frac{\lambda}{\lambda_z} = \frac{\lambda}{\lambda_{z0}} \left(1 - \frac{\lambda_{z0}^2}{2\pi^2 h w} \frac{p}{1 + p^2} \right) \quad (11-21)$$

and

$$\alpha_z \lambda = \frac{\lambda \lambda_{z0}}{\pi h w} \frac{1}{1 + p^2} \quad (11-22)$$

where

$$p = \frac{2}{\pi} \left[\ln \left(\csc \frac{\pi \delta}{2w} \right) + \ln \left(1.526 \frac{h}{\delta} \right) \right] \quad (11-23)$$

and λ_{z0} is the guide wavelength, that is,

$$\lambda_{z0} = \lambda / \sqrt{1 - (\lambda/2h)^2} \quad (11-24)$$

in the unperturbed waveguide ($\delta = 0$). Zero ground-plane thickness is assumed. Although λ/λ_c and $\alpha_z\lambda$ are not strictly separable, λ/λ_c is controlled primarily by variations in h (through λ_{c0}) and $\alpha_z\lambda$ by variations in δ . For operation at frequencies near cutoff, you should use the exact solution (also in Goldstone and Oliner¹⁷) instead of the perturbation form given here.

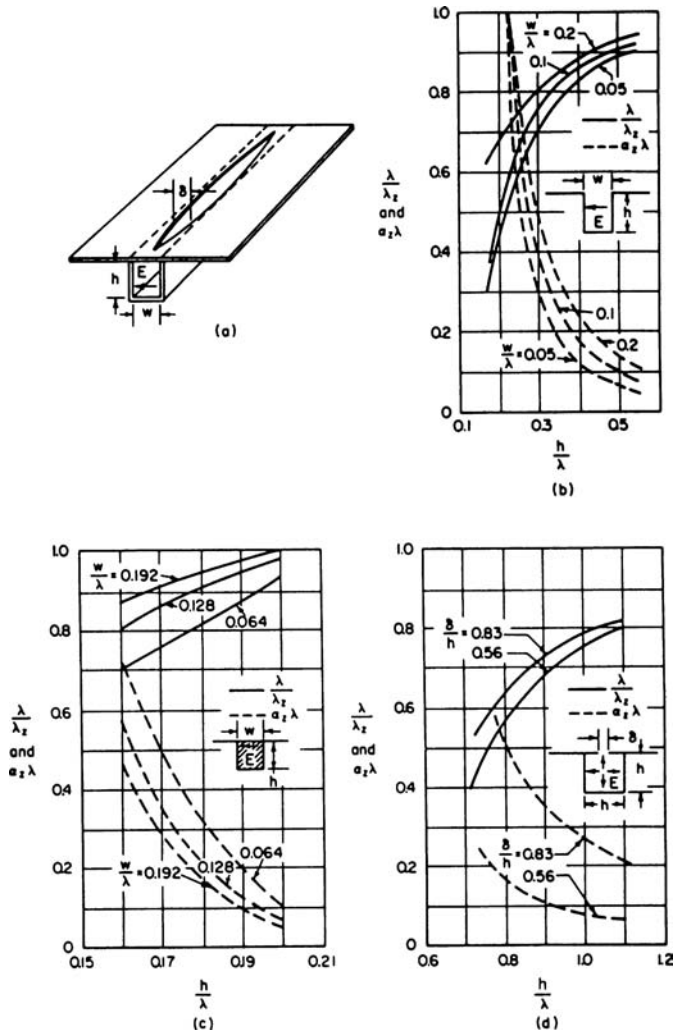


FIGURE 11-5 Several examples of leaky-wave antennas based on long slits in rectangular waveguide. (a) Narrow slit in the narrow wall, shown with an exaggerated taper. The remaining figures show the relative guide wavelength λ/λ_c and the relative leakage constant $\alpha_z\lambda$ for (b) the channel-guide antenna, (c) the dielectric-filled channel, for $\epsilon_r = 2.56$, and (d) a narrow slit in square guide. The dominant mode excites the slit in the first three cases, while the TM_{11} mode excites the slit in the last case (after Goldstone and Oliner¹⁷ © IRE (now IEEE) 1959 and Hines et al¹⁸ © IRE 1953).

When the rectangular waveguide is dielectric-filled, Eqs. 11-21 to 11-24 must be modified appropriately; such expressions are given in Walter,¹² on pages 189 and 190.

Although expressions corresponding to Eqs. 11-21 through 11-24 are also available in Goldstone and Oliner¹⁷ for the structures in Figure 11-5*b* and *d*, and in Walter¹² for that in Figure 11-5*c*, results for these structures are displayed graphically in Figure 11-5 to illustrate how the numerical values for normalized β and α vary with frequency. Theoretical expressions for these long slit structures have also been derived by Rumsey¹⁹ and by Hines et al¹⁸ using a variational approach. Modifications of the theoretical expressions given in Goldstone and Oliner¹⁷ for these structures when dielectric loading is present are contained in Chapter 5 of Walter.¹²

It should be recognized that the slit fields in the structures in Figure 11-5 *a*, *b*, and *c* have $E_z = 0$, and that the transverse electric field in the aperture is equivalent to a longitudinal (z -directed) magnetic current. The resulting radiation pattern is horizontally polarized. In contrast, the slit in the structure in Figure 11-5*d* has $H_z = 0$ and primarily constitutes a longitudinal electric current, so that the antenna radiates with vertical polarization.

The configurations in Figure 11-5*b* and *c*, for which the slit has been opened up to the full guide width (w), are generally referred to as *channel-guide antennas*.^{20,21} Because the slit is so wide, however, the leakage constant can become large rather easily. Even in the structure in Figure 11-5*a*, however, the slit cuts directly across the electric field lines, corresponding to the logarithmic dependence seen in Eq. 11-23, so that the leakage rate will never be very small. If a leaky-wave antenna with a very narrow beam is required, therefore, you need to go to a structure like that containing the series of closely spaced holes described later in this section.

Measurements have been taken on all these structures, and the agreement with theory has been excellent. However, most of these measurements have been on untapered slits. Measurements on tapered-slit antennas have been reported, e.g., in Hines et al¹⁸ and Rotman.²¹ For the antenna in Figure 11-5*a*, good control over the sidelobes has been achieved,¹⁸ a typical result being that a Gaussian amplitude distribution produces low sidelobes over an almost 2:1 frequency range in which the beam swings from 38 to 18° off end fire.

Many years later, a modified long-slot structure was proposed and analyzed²² that provides radiation patterns with very low sidelobes within a relatively short length and with high efficiency. The long slot is located on the top wall of rectangular waveguide, and it moves from the centerline of the top wall out to near the side wall and back to the centerline in a very strong, but specified, curved fashion. The designers call this variation their *meander* contour. Numerical results were provided, though no experimental verification was given.

Long Slits in Circular Waveguide

The radiation properties of leaky-wave antennas based on long slits in circular waveguide are qualitatively similar to those for long slits in rectangular waveguide, as discussed earlier. Three independent theoretical analyses appear to be available. The first two, by Harrington²³ and by Rumsey,¹⁹ employ a variational approach, and the authors present accurate results that agree very well with their independently obtained measurements. The third analysis, by Goldstone and Oliner,²⁴ uses a transverse resonance approach in the radial direction and develops expressions for the slit using radial transmission-line theory. The authors obtain explicit expressions for the phase and leakage constants in a relatively simple, explicit form, in contrast to the other two theories, and these expressions are also much simpler to compute from. The results also agreed well with Harrington's measurements.²³

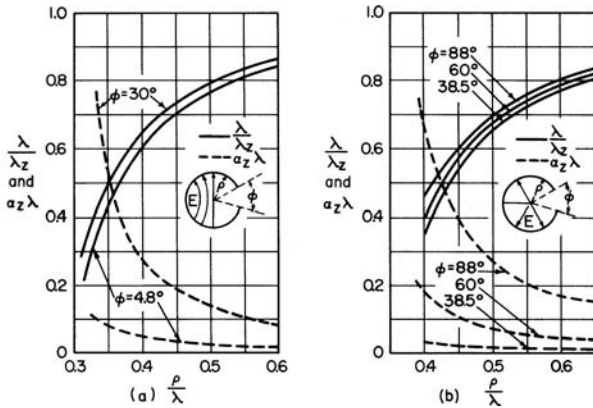


FIGURE 11-6 Long slits in circular waveguide: (a) TE_{10} mode excitation; (b) TM_{01} mode excitation (after Harrington²³ and Goldstone and Oliner²⁴ © IRE (now IEEE) 1961)

Numerical results for air-filled circular waveguides supporting the TE_{11} mode and the TM_{01} mode, to furnish opposite types of polarization, are presented in Figure 11-6a and b, respectively. Explicit expressions for air-filled guides are given in Goldstone and Oliner,²⁴ and the modifications in those expressions for the dielectric-filled case are included in Chapter 5 of Walter.¹²

Closely Spaced Holes in Rectangular Waveguide

A problem relating to the structures in Figure 11-5a–c is that the opening directly disrupts the current lines, so you cannot, with such a geometry, obtain a very narrow radiated beam. A way to surmount this difficulty is to replace the long slit (or the channel-guide geometry) with a series of closely spaced small holes, as shown in Figure 11-7. Then the current lines are simply pushed aside by the holes, and they can go around them. Since the holes perturb the initially closed guide much less than the long slit does, the resulting value of α is much smaller. Finally, since the holes are closely spaced, the structure is quasi-uniform and the design principles in Section 11.2 apply.

The antenna employing a series of closely spaced round holes was proposed and measured at the Ohio State University,²⁵ and it became known as the OSU “holey guide.” By varying the diameter d of the holes and, to a lesser extent, the hole spacing s and the guide width w , a very large range in the value of $\alpha\lambda_z$ was found experimentally. This structure was also analyzed by Goldstone and Oliner,¹⁷ by employing small aperture procedures for the series of holes, together with a transverse resonance approach. Using a perturbation form for the

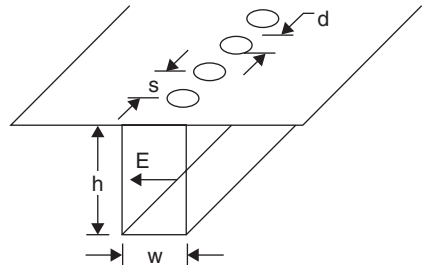


FIGURE 11-7 Quasi-uniform aperture in rectangular waveguide to permit narrow radiated beams; closely spaced round holes in the narrow wall (“holey guide”) (after Hines and Upson²⁵)

result, they obtained the following simple expressions (using the notation in Eqs. 11-21 and 11-22):

$$\frac{\lambda}{\lambda_z} = \frac{\lambda}{\lambda_{z0}} \left[1 + \frac{\lambda_{z0}^2}{4\pi h^2} X' \right] \quad (11-25)$$

and

$$\alpha_z \lambda = \frac{\lambda \lambda_{z0} R'}{2h^2} \quad (11-26)$$

where

$$X' = \frac{B'}{(G')^2 + (B')^2} \quad R' = \frac{G'}{(G')^2 + (B')^2} \quad (11-27)$$

$$G' = \pi w / 2h \quad (11-28)$$

$$B' = 6hws / \pi d^3 \quad (11-29)$$

$$\lambda_{z0} = \lambda / \sqrt{1 - (\lambda / 2h)^2} \quad (11-30)$$

Calculations from these expressions agreed very well with measurements made both at the Ohio State University (Walter,¹² Chapter 5) and the Polytechnic Institute of Brooklyn.¹⁷ When the frequency of operation is near to cutoff, these perturbation expressions become inaccurate, and we should instead solve the transverse resonance relation¹⁷ exactly.

Array of Closely Spaced Wide Transverse Strips

This array of transverse strips, sometimes called an *inductive-grid antenna* and due to R. C. Honey,²⁶ is shown in Figure 11-8a. The antenna consists of a parallel-plate waveguide operated in its first higher-order (TE₁) mode, with its upper plate composed of an array of closely spaced transverse strips, and fed from a reflector arrangement so as to fill the space with the field having the polarization shown in Figure 11-8a. The upper plate can be photoetched on a thin laminate and then be supported by polyfoam, or it can consist of a grid of transverse round wires.

The structure was analyzed²⁶ by using the transverse resonance method, which yielded simple and accurate expressions for the λ/λ_z and $\alpha_z \lambda$ values. From these expressions we can compute the design curves presented in Figure 11-8b and c. The design procedure is to first select the desired λ/λ_z and $\alpha_z \lambda$ and then to read from the curves in Figure 11-8b the corresponding abscissa value. The curves in Figure 11-8c then yield the value of d for the design wavelength corresponding to the abscissa value obtained from Figure 11-8b. Now, in a design for low sidelobes in some specified fashion (see Section 11.2), $\alpha_z \lambda$ must vary from point to point in a tapered fashion along the longitudinal direction, while λ/λ_z must remain the same at each point. The plot in Figure 11-8b then tells us how c can be varied, by changing the strip width t , to obtain the desired different values of $\alpha_z \lambda$ while trying

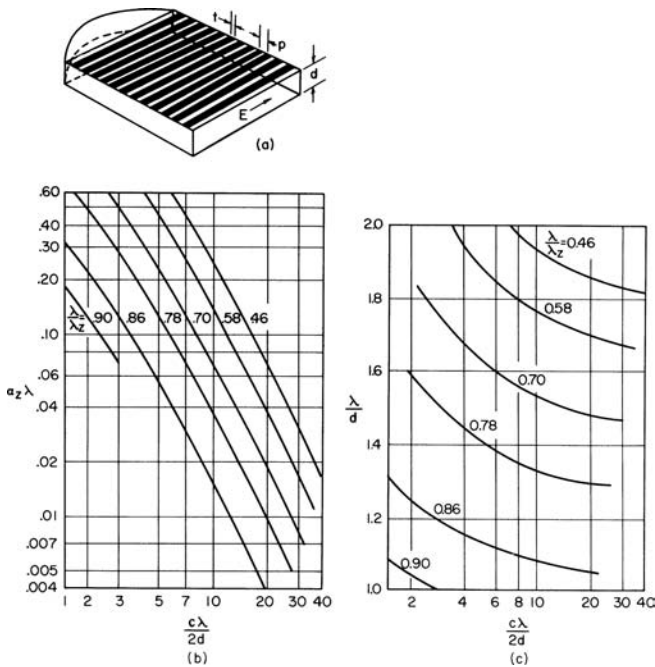


FIGURE 11-8 Inductive-grid antenna comprised of a parallel-plate guide fed in its first higher-order TE mode and with its upper plate consisting of a series of closely spaced wide transverse strips: (a) structure; (b) relative leakage constant with the normalized guide wavelength as a parameter; (c) a plot to aid in the design procedure (see text). Quantity c , which appears in the abscissas of parts (b) and (c), is defined as $c = 2\pi d/p \ln [\csc (\pi/2p)]$ (after Honey²⁶ © IRE (now IEEE) 1959).

to maintain λ/λ_z constant. However, λ/λ_z will change somewhat as t is varied, since it is not independent of t , and the plot in Figure 11-8c then indicates how d can be modified to change λ/λ_z back to the desired constant value.

In his final design, Honey²⁶ found that it was necessary to flex the bottom plate slightly along the longitudinal direction, and he built his structure accordingly. He was also meticulous with respect to both the accuracy of his theory and the details of the structure to be measured. As a result, the correspondence between his theoretical and measured radiation patterns was *remarkably good*, down to almost -40 dB, as may be observed in Figure 11-9.

11.6 SPECIFIC STRUCTURES BASED ON PERIODIC OPEN WAVEGUIDES

The design principles for leaky-wave antennas based on periodic open waveguides are presented in Section 11.3. The important points to recall are that the basic open waveguide supports a slow wave, which does not radiate, and that the period of the structural modulation is selected relative to the wavelength so that the $n = -1$ space harmonic, and only that

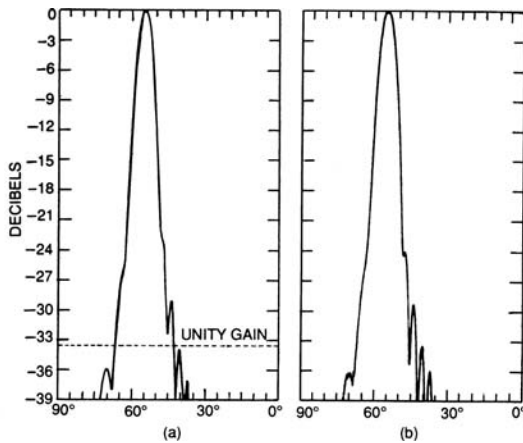


FIGURE 11-9 Theoretical (a) and measured (b) radiation patterns for the inductive-grid antenna shown in Figure 11-8a. The accurate theoretical design and the carefully fabricated experimental structure both took into account the taper for low sidelobes. The agreement between the two patterns is seen to be remarkable (after Honey²⁶ © IRE (now IEEE) 1959).

one, radiates the power. In contrast, the periodic structures discussed in Section 11.5, such as the closely spaced series of round holes in rectangular waveguide, are based on a fast wave and are quasi-uniform so that only the $n = 0$ space harmonic radiates. Furthermore, the beam radiated from the periodically modulated slow-wave structures may be scanned throughout most of the backward quadrant and into part of the forward quadrant, whereas the beam radiated from the uniform or quasi-uniform structures is restricted to the forward quadrant only.

The two most common open waveguides that support dominant modes and which serve as the basis for periodic leaky-wave antennas are rectangular dielectric rods, with or without ground plane, and microstrip line. The class of dielectric rods includes a variety of known waveguides, such as dielectric image guide, insular guide, inset guide, and so on. Since the leaky wave is fast and the basic surface wave (or microstrip dominant mode) is slow, we must be careful about the feed arrangement to make sure that little spurious radiation is introduced.

Many of these periodic leaky-wave antennas have been known (measured and used) for some years, but accurate theories for β and α , suitable for careful design purposes, have become available only recently, and only for some structures.

Early Structures

A few pioneering examples in this class were proposed and studied as far back as the late 1950s, but the ideas behind them were not pursued then. When they did reemerge, a decade or two later and in a somewhat different form, most people did not recognize the relationship with the past. Two examples of this early novel thinking are presented here.

The first structure is based on the original dielectric image guide, due to D. D. King,²⁷ which consists of the top half of a round dielectric rod placed on a ground plane.

The fields of the dominant mode on the rod (the TE_{11} mode) extended substantially into the air region transversely, as did the current lines on the ground plane. In the leaky-wave antenna configuration²⁸ based on this guide, a two-dimensional (2D) array of slots was cut into the ground plane on each side of the rod. In this way the (loosely bound) surface wave guided by the rod excites each *row* of slots in phase, with an illumination corresponding to the transverse decay rate of the field, and each *column* of slots in accordance with the propagation wave number β . Depending on the type, orientation, and spacing of the slots, a variety of radiation patterns can be obtained. A summary of the possibilities, together with sketches of some slot configurations, appear on pages 16-33 and 16-34 of Zucker.¹³ This antenna was proposed for application to millimeter waves, but its concept was not reintroduced until interest in millimeter waves was revived 15 years or so later.

The second antenna structure was called the *sandwich-wire antenna*, and its configuration was an outgrowth of stripline. There are two versions. In the first configuration,²⁹ the top and bottom plates and the center strip of stripline are each reduced to wires, and the center wire is then snaked along its length in periodic fashion. The periodicity is selected so that the $n = -1$ space harmonic can radiate. A second configuration³⁰ looks more like a suspended microstrip line that is cavity-backed, with a strip mounted on a cavity-backed thin dielectric layer, to support the strip, and with the strip undulating sinusoidally back and forth along the dielectric layer. Again, the $n = -1$ space harmonic radiates, and measurements were made to determine the performance properties. Diagrams of the structures and curves for some measurements may be found on pages 16-36 and 16-38 of Zucker.¹³ A limitation on the usefulness of this antenna approach is that dimensional changes that affect β affect α as well. For example, a good approximation for the relative phase velocity is obtained by assuming that the wave velocity along the undulating strip is that of light. Thus if the amplitude of the undulation is increased, the value of β will be increased. However, the value of α is changed most easily by varying that same amplitude.

Periodic Dielectric Waveguides

Periodic dielectric waveguides are uniform dielectric waveguides with a periodic surface perturbation. Several types of uniform dielectric waveguides are shown in cross section in Figure 11-10a through e. Rectangular shapes are preferred for antennas of this class. The type in Figure 11-10a is a simple rectangular rod of rectangular shape; the type in Figure 11-10b is a rectangular form of dielectric image guide; and the type in Figure 11-10c, the insular guide, has an extra dielectric layer on the ground plane so as to reduce the ground-plane losses. The types in Figure 11-10d and e, the trapped image guide³¹ and the inset guide,³² respectively, lend themselves to a flush-mounted arrangement and reduce radiation losses from bends. The two most common dielectric waveguides are the types in Figure 11-10a and b.

The most common periodic modulation methods are a grating of grooves, a grating of metal strips, and, to a lesser extent, a series of metal disks. These perturbations are ordinarily placed on the top surface (the wide dimension) of these guides, but they also may be placed on the sides,

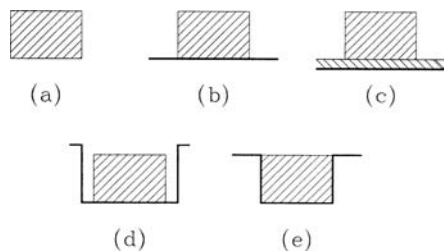


FIGURE 11-10 Cross sections of several types of uniform dielectric waveguides: (a) rectangular rod of rectangular shape; (b) rectangular form of dielectric image guide; (c) the insular guide, with an extra dielectric layer of lower ϵ_r on the ground plane; (d) trapped image guide; (e) inset guide

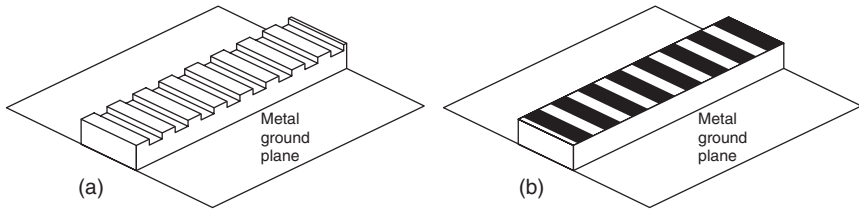


FIGURE 11-11 Dielectric image guide with (a) a grating of grooves and (b) a grating of metal strips

when accessible. Dielectric image-guide antennas with a grating of grooves and with a grating of metal strips appear in Figure 11-11a and b, respectively. These gratings are shown as uniform; in practice, the gratings would be tapered so that the groove depths and the metal strip widths would be very small at the beginning and the end in each case.

The antennas in Figure 11-11a and b were first proposed around 1960, but systematic investigations of their behavior, both experimental and theoretical, were not carried out until the late 1970s. Most of the early studies were conducted by the army at Ft. Monmouth, on grooved antennas^{33–35} and on antennas with metal strips.^{36–38} Experimental studies on the metal-strip antennas were also performed at the University of Illinois^{39–41} during this period on dielectric image guide, and by Itoh and Adelseck⁴² on trapped image guide. One important conclusion from the Ft. Monmouth studies was that in practice there is an upper limit to the value of leakage constant obtainable with grooved gratings; antennas employing grooved gratings were therefore limited to narrow radiated beams. It was found that the metal-strip gratings were more versatile, permitting both wide and narrow beams.

Studies also were concerned with how these antennas could be fed without causing spurious radiation at the feed.^{40,43} Solbach⁴³ utilizes slots cut into the ground plane under the dielectric rectangular rod to effect a smooth transfer of power from the feedline.

The radiation pattern in the cross-plane depends on the width of the dielectric structure. The waveguide is excited in its dominant mode, and its width is usually chosen to be comparatively small so as to avoid excitation of higher-order modes. Alternatively, if you wish to narrow the beam in the cross-plane, the dielectric width may be made large, but then single-mode excitation must be ensured by an appropriate feed arrangement. Another technique⁴¹ is shown in Figure 11-12. The grating structure is embedded in a rectangular metal trough, which is then attached to a flared-horn configuration; design details are included in Trinh et al.⁴¹

Design Theory for Wide Periodic Dielectric Antennas

Starting in the late 1970s accurate theoretical analyses became available that permit the systematic design of this class of leaky-wave antennas. For antennas employing a *grating of grooves*, thorough and detailed studies were performed by Schwering and Peng,^{7,33,34} based in part on earlier analytical work by Peng and Tamir.^{44,45} The two best sources for systematic design information are Schwering and Peng⁷ and Schwering and Oliner.⁴⁶ Corresponding, but less thorough,

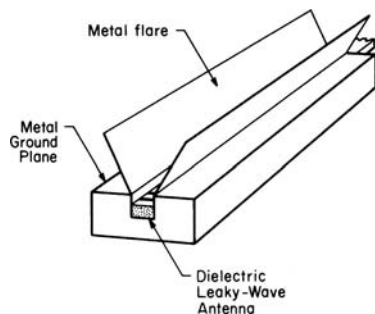


FIGURE 11-12 The use of a flared horn in conjunction with the structure in Figure 11-11b to narrow the beamwidth in the cross-plane (after Trinh et al.⁴¹ © IEEE 1981)

design information on *metal-strip-grating antennas* comes mostly from work by Guglielmi and Oliner,⁴⁷ based on earlier analyses by them⁴⁸⁻⁵⁰ of scattering by metal strip gratings on a dielectric substrate. Another accurate method of analysis for these antennas was presented by Encinar.⁵¹ Many of the general conclusions appropriate to grooved antennas, however, apply as well to those with metal-strip gratings. The detailed expressions for β/k_0 and α/k_0 are different, of course.

The theoretical design information referred to earlier is applicable directly to dielectric image guides, i.e., structures for which a ground plane is present under the rectangular dielectric layer. The procedure is readily extendable to structures without a ground plane, however, by some suitable, basically straightforward modifications. The numerical values presented in Schwering and Peng⁷ and Schwering and Oliner⁴⁶ assume that the dielectric material has an $\epsilon_r = 12$, corresponding to Si or GaAs, so that the antenna performance can be controlled, if desired, by semiconductor devices.

For the antenna employing a *grating of grooves*, the groove depth must be chosen to lie within a certain range if we wish to optimize the leakage constant α . The reason for this can be understood physically in a simple way. Consider the structure shown in the inset in Figure 11-13, where the height of the uniform dielectric region is h and the groove depth is t . Suppose that we maintain the sum of $h + t$ constant and we increase the groove depth t . Although the total antenna height is fixed, the effective dielectric constant ϵ_{eff} of the structure is decreased as t is increased, because the groove region is

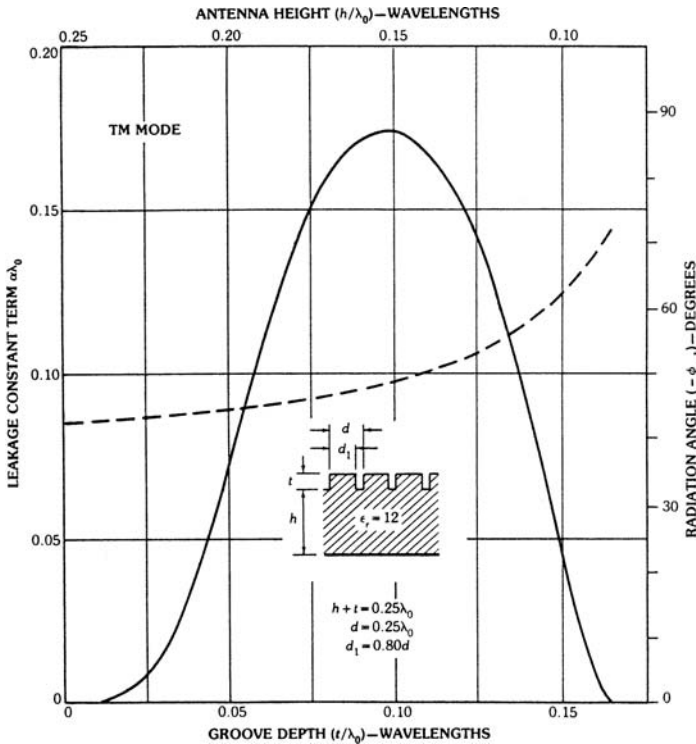


FIGURE 11-13 Calculations for the normalized leakage constant and the radiation angle as a function of groove depth for the dielectric image guide with a grating of grooves shown in Figure 11-11a (after Schwering and Peng⁷ © IEEE 1983)

now partly air-filled. The value of ϵ_{eff} , combined with the height $h + t$, may be viewed as an “effective height” and is an important design parameter. When the effective height is small, most of the guided energy travels in the air region above the antenna, and the grooves would cause little radiation. When the effective height is large, on the other hand, the energy is confined primarily to the interior of the antenna, and again the grooves will have little effect. An intermediate effective height thus exists for which the energy density in the grooved region reaches a peak value. We would therefore expect that α could be maximized by an optimal combination of groove depth and effective height. The curve of $\alpha\lambda_0$ versus t/λ shown in Figure 11-13 illustrates precisely such behavior for the structure treated there. Similar qualitative reasoning applies to other periodically modulated open dielectric structures.

Design Theory for Narrow Periodic Dielectric Antennas

The theoretical results referred to in the above subsection make the assumption that the dielectric structures have infinite width w . It has been found, however, that they apply quite accurately to “wide” structures, for which $w > \lambda_0 / \sqrt{\epsilon_{\text{eff}} - 1}$. If $\epsilon_{\text{eff}} = 2$, for example, these results are applicable to antennas for which $w > \lambda_0$. For $\epsilon_{\text{eff}} = 6$, on the other hand, w need satisfy only $w > 0.45\lambda_0$. Thus the theory for “infinite” width can actually be applied with good accuracy to structures that are fairly narrow. When the antennas are narrower than the criterion just mentioned will allow, a correction scheme is available that yields accurate results for them as well.

For antennas of narrow width ($w < \lambda_0 / \sqrt{\epsilon_{\text{eff}} - 1}$), the value of β can be derived in a simple way with good accuracy by using the EDC (equivalent dielectric constant) procedure, but α can no longer be approximated by that of an infinitely wide antenna with the same dielectric constant as the finite antenna. As w decreases, the phase velocity of the leaky mode increases, so that an increasing portion of the guided energy now travels in the air regions on both sides in the neighborhood of the perturbing mechanism, whether grooves or metal strips. As a result, the leakage constant α of the narrower antenna becomes smaller.

A simple procedure has been developed⁵² that employs the EDC method to replace the antenna of finite width by an equivalent antenna of infinite width, but with a lower effective dielectric constant. As a result, a lower value of α is obtained. This procedure has produced numerical values for α and β that have agreed very well with measured results for a quite narrow antenna, with $\epsilon_r = 16$ and $w = 1.3$ mm, in the frequency range from 30 to 36 GHz.⁵² Some details regarding this procedure may also be found in Schwering and Oliner,⁴⁶ pages 17-64 to 17-68.

Periodic Leaky-Wave Antennas Based on Microstrip Line

A wide variety of possible traveling-wave periodic array antennas can be achieved by employing microstrip line. As examples, we can employ a series of resonant patch antennas connected by the microstrip line, as seen in Figure 11-14a, or a series of array elements coupled by proximity to the

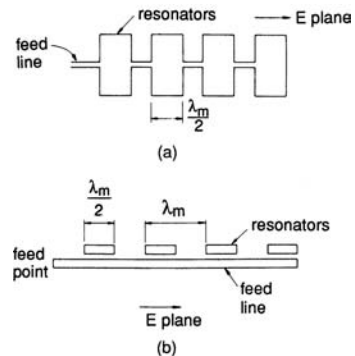


FIGURE 11-14 Top views of traveling-wave periodic array antennas based on microstrip line: (a) series of resonant patch antennas connected by microstrip line; (b) series of elementary radiators, resonant or not, coupled by proximity to the microstrip line (after James et al⁵³)

microstrip line, as shown in Figure 11-14*b*. Top views of the structures are presented. Another wide class of possibilities is illustrated in Figure 11-15*a* through *c*, and these involve periodic meanderings of the microstrip line strip itself. These last structures are reminiscent of the sandwich-wire antenna^{29,30} described earlier in the subsection “Early Structures.” These and other traveling-wave arrays based on microstrip line are discussed in detail in the two comprehensive books^{53,54} on microstrip antennas by James et al and by James and Hall, respectively. The original references for the antennas in Figure 11-14*a* and *b* are Derneryd⁵⁵ and Cashen et al,⁵⁶ respectively.

No theory is available for most of these antennas. For the few structures, such as the one shown in Figure 11-14*a*, for which some theory is available, the theory is of the type used to describe the behavior of slot arrays. All the thinking, in fact, parallels that employed for slot arrays. For example, the array elements are usually assumed to be individually resonant, although there is no reason why they need to be. The theory then treats each element as loading the line individually, instead of viewing the structure in leaky-wave fashion. (See the discussion in Section 11.1 on the relation between leaky-wave antennas and slot arrays.) Furthermore, in many cases the arrays are designed to be resonant (standing wave rather than traveling wave). The array in Figure 11-14*a* is specified as a resonant one, but an interesting traveling-wave modification has been reported.⁵⁷ This modification contains additional phase shift between successive elements to reduce the frequency change needed to cover a given range of scan angle.

These same structures can be transformed into standard leaky-wave antennas by simply making the individual radiating elements nonresonant (so each loads the line less strongly) and changing the spacing between elements to produce a traveling-wave rather than a standing-wave array. Then the structure would lend itself to a leaky-wave analysis, although such an analysis has not yet appeared.

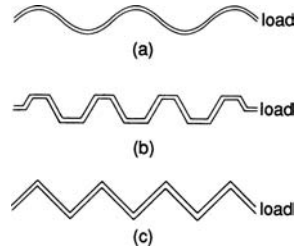


FIGURE 11-15 Examples of periodic leaky-wave antennas based on microstrip line where the microstrip line itself is meandered periodically: (a) sinusoidal, (b) trapezoidal, and (c) zigzag (after James et al⁵³)

Periodic Arrays of Microstrip Patches or Dielectric Resonators Fed by Open Dielectric Waveguides

Since microstrip line becomes increasingly lossy as the frequency is raised, several investigators have built and measured antennas designed for use at millimeter wavelengths in which the microstrip line is replaced by a dielectric waveguide as the transmission line that feeds the periodic radiating elements. The viewpoint is that the feedline can be considered separately from the radiating elements, so that a less lossy feedline can be employed at the higher frequencies. The resulting antennas are indeed less lossy, but they sometimes become hybrid in character.

An example of such antennas is shown in Figure 11-16. The radiating elements here are microstrip patch resonators, and they are fed by a low-loss open dielectric waveguide in the form of insular guide. The fringing fields of the dielectric guide excite the microwave patches. It is interesting that this mechanism is the same as the one used many years ago²⁸ with dielectric image line, discussed earlier in the subsection “Early Structures.” In addition to low loss, this antenna has the advantage of simplicity. In an experimental design^{58,59} the distance between the metal patches and the dielectric guide is varied along the length to produce a taper to control the sidelobe level, as seen in Figure 11-16*b*. Nevertheless, the design was purely empirical, since no theory was available, and the resulting radiation-pattern performance was only mediocre.

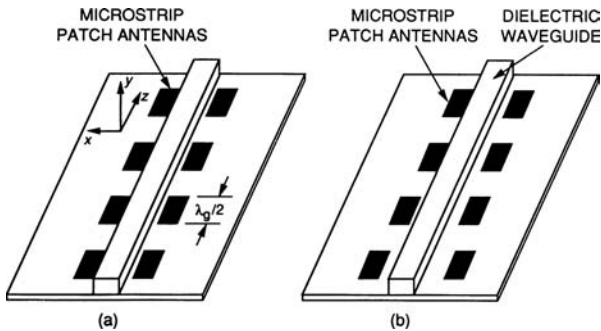


FIGURE 11-16 A hybrid form of periodic leaky-wave antenna in which microstrip patch resonator arrays are fed by a dielectric image guide: (a) basic configuration; (b) configuration including a taper to control the sidelobes (after James and Henderson⁵⁹)

At about the same time, another study⁶⁰ replaced the microstrip patches by dielectric resonators, which may take the form of small rectangular or cylindrical dielectric blocks. The same type of open dielectric waveguide served as the feed transmission line. Again, however, the design was only empirical, and the pattern performance was therefore only marginal.

11.7 SPECIFIC STRUCTURES BASED ON UNIFORM OPEN WAVEGUIDES

As indicated in Section 11.5, almost all the early uniform leaky-wave antennas were based on initially closed waveguides. Those antennas were conceptually simple, and they worked very well. It was only much later, in response to requirements at millimeter wavelengths, that thought was given to leaky-wave antennas based on uniform *open* waveguides. There exists one notable exception, and that case is discussed later in the subsection “Early Structure: Asymmetrical Trough Waveguide Antenna.”

There were two reasons why new types of leaky-wave antennas were sought for the millimeter-wave region. The first relates to the small wavelengths in this region; to minimize fabrication difficulties, the new antennas had to be *simple* in configuration, and *uniform* guiding structures satisfied this condition. The second reason was that the usual waveguides had higher *loss* at these higher frequencies; as a result, the new antennas were generally based on new lower-loss waveguides that had been studied specifically for application to the millimeter-wave region. These waveguides were *open* so that part of the field extended into the air region outside, thereby reducing the energy density and the loss. The principal waveguides in this category were nonradiative dielectric (NRD) guide and groove guide.

Because of the open nature of these waveguides, new mechanisms had to be found to produce the leakage. A physical cut is not meaningful because the structure is already open. Three main mechanisms were employed: (1) introducing asymmetry in the structure so that a radiating component of field is created, (2) foreshortening some dimension in the cross section, and (3) using a higher-order mode that is leaky in itself, rather than the bound dominant mode. The mechanism most commonly employed then and today is asymmetry.

Many geometric arrangements can be devised that satisfy the preceding requirements and make use of one of the leakage mechanisms listed, but we must remember that the resulting structures also must be analyzable so that the antennas can be designed systematically

and not simply on an empirical basis, as is true for some of the more novel periodic open leaky-wave antennas discussed in the preceding section. All the uniform open antennas described in this section have been analyzed accurately, and theoretical expressions for their design are available in the literature. Most of the antennas discussed here are due to Oliner, as part of a systematic cooperative investigation with principal colleagues Lampariello of Italy and Shigesawa of Japan. Summaries of their contributions appear in two books^{46,61} and in a review article⁶²; details are contained in individual papers referenced later and in a comprehensive two-volume report.⁶³

Early Structure: Asymmetrical Trough Waveguide Antenna

The first open waveguide that was made leaky by introducing asymmetry in the cross section was the trough waveguide. The antenna structure, shown in Figure 11-17 in full view and in cross section, was invented and measured by Rotman⁶⁴ and analyzed by Oliner⁶⁵ about 40 years ago. It was a practical antenna, widely used, and is still useful today. It appears to be the only early example of the class of uniform leaky-wave antennas based on open waveguides.

The trough waveguide itself, when operated as a nonradiating transmission structure, is symmetrical about the center fin and is derived from symmetrical strip transmission line by placing a short-circuiting plate at its midplane. The dominant mode in the trough waveguide is therefore identical with the first higher-order mode in stripline. Trough waveguide

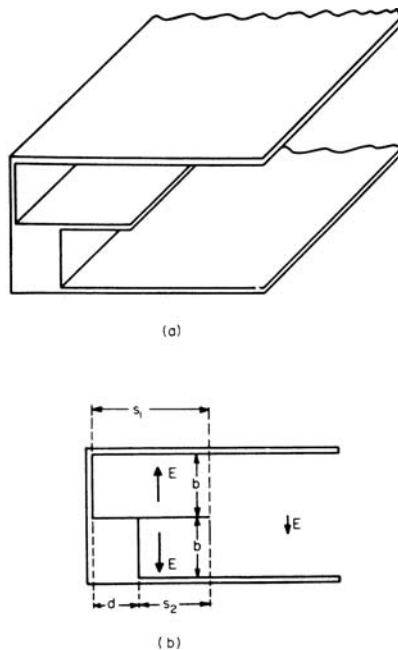


FIGURE 11-17 The uniformly asymmetrical trough waveguide antenna: (a) full view and (b) cross section for zero-thickness center fin showing the electric field orientations (after Rotman and Oliner⁶⁵ © IRE (now IEEE) 1959)

therefore combines the mechanical simplicity of a stripline with the frequency characteristics of a waveguide, and its bandwidth for single-mode propagation is greater than that for rectangular waveguide by about 50 percent. It also can be coupled smoothly to a coaxial line, a feature that makes it convenient for use at lower frequencies.

Despite the fact that trough waveguide is open on one side, it is nonradiating when the structure is symmetrical. The introduction of asymmetry, however, will produce radiation in a leaky-wave fashion. In fact, one virtue of this type of leaky-wave antenna is the simple means by which radiation can be controlled. The asymmetry can be varied, for example, by placing a metal insert in one of the halves of the line, as shown in Figure 11-17, and adjusting its thickness d .

A theoretical analysis⁶⁵ employing transverse resonance together with a perturbation treatment yields the following simple expressions for the leaky-wave propagation characteristics:

$$\frac{\lambda}{\lambda_g} = \frac{\lambda}{\lambda_{g0}} - \frac{\lambda}{2\pi} \frac{\lambda_{g0}}{\lambda_{c0}} \text{Re}(\Delta\kappa) \quad (11-31)$$

and

$$\alpha = \frac{\lambda_{g0}}{\lambda_{c0}} \text{Im}(\Delta\kappa) \quad (11-32)$$

where

$$\text{Re}(\Delta\kappa) = \frac{\pi d}{4L_1L_2} \left(1 - \frac{d}{L_1 + L_2} \right) \quad (11-33)$$

and

$$\text{Im}(\Delta\kappa) = \frac{(\pi d)^2}{8L_1L_2(L_1 + L_2)} \quad (11-34)$$

with

$$L_1 = s_1 + \frac{2b}{\pi} \ln 2 \quad (11-35)$$

and

$$L_2 = s_2 + \frac{2b}{\pi} \ln 2$$

The structural dimensions s_1 , s_2 , d , and b are indicated in Figure 11-17; λ_{c0} and λ_{g0} are the cutoff wavelength and guide wavelength respectively of the symmetrical and therefore nonradiating trough waveguide.

These expressions have been found to give rather good agreement with measured values,⁶⁵ especially for narrow radiated beams. For wider beams it is better to use the transverse resonance expression itself (without the perturbation simplifications), which is also given and derived in Rotman and Oliner.⁶⁵ Numerical data for various parameter combinations, and corrections to be made for center fins of appreciable thickness, are also found in this reference.

Foreshortened-Top NRD Guide Antenna

Nonradiative dielectric (NRD) guide is a low-loss open waveguide for millimeter waves that was first proposed and described in 1981.⁶⁶ It is a modification of H guide where the

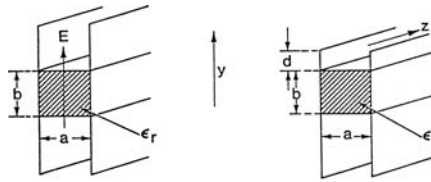


FIGURE 11-18 The nonradiative dielectric (NRD) guide, on the left, and the foreshortened-top leaky-wave antenna based on it, on the right (after Sanchez and Oliner⁶⁷ © IEEE 1987)

spacing between the metal plates is less than $\lambda_0/2$ so that all junctions and discontinuities that maintain symmetry become purely reactive instead of possessing radiative content. The waveguide structure is shown on the left-hand side of Figure 11-18. The dielectric material in the center portion confines the main part of the field, and the field decays exponentially in the vertical direction in the air region away from the dielectric-air interfaces. Two new leaky-wave antennas are based on this waveguide: the foreshortened-top NRD guide antenna and the asymmetrical NRD guide antenna. The former antenna is discussed now, and the latter one is treated in the next subsection.

When the vertical metal plates in the NRD guide are sufficiently long, as shown on the left-hand side of Figure 11-18, the dominant mode field is effectively completely bound, since the field has decayed to negligible values as it reaches the upper and lower open ends. If the upper portion of the plates is foreshortened, as seen on the right-hand side of Figure 11-18, the field is still finite at the upper open end (but negligible at the lower open end). A traveling-wave field of finite amplitude then exists along the length of the upper open end, and, if the dominant NRD guide mode is fast (it can be fast or slow depending on the frequency), power will be radiated away at an angle from this open end.

An accurate theory was developed by Sanchez and Oliner,⁶⁷ in which an accurate transverse equivalent network was developed for the cross section of the antenna, and the dispersion relation for the values of α and β was obtained from the resonance of this network. All the elements of this dispersion relation are in closed form, thus permitting easy calculation. The leakage constant α is determined simply by the amount of foreshortening, measured by d , and β is essentially unaffected by changes in d unless d becomes very small. Thus α and β can be adjusted independently, to a great extent, which is very desirable because the procedure for sidelobe control is then simplified.

Careful measurements were taken by Yoneyama⁶⁸ at 50 GHz and by Han et al⁶⁹ in the vicinity of 10 GHz on a scaled structure. Excellent agreement between measurement and theory was found over the range of parameter values examined, and results for one case are given on Figure 11-19. You also can see that α can be varied over an extremely wide range, permitting narrow beams or wide beams, simply by altering the value of d .

The amplitude distribution in the radiating aperture can be controlled by varying the distance d as a function of the longitudinal coordinate z . This may be accomplished either by appropriately shaping the upper edge of the metal plates or by slightly curving the dielectric strip so that its distance from the edge of the plates varies in a prescribed fashion with z ; the edge in this case would be straight. The latter procedure leads to a very simple and easy-to-build antenna configuration, which is indicated in Figure 11-20. Furthermore, the antenna is directly compatible with transmit and receive circuits designed in NRD guide technology. The antenna radiates with vertical polarization in the principal plane.

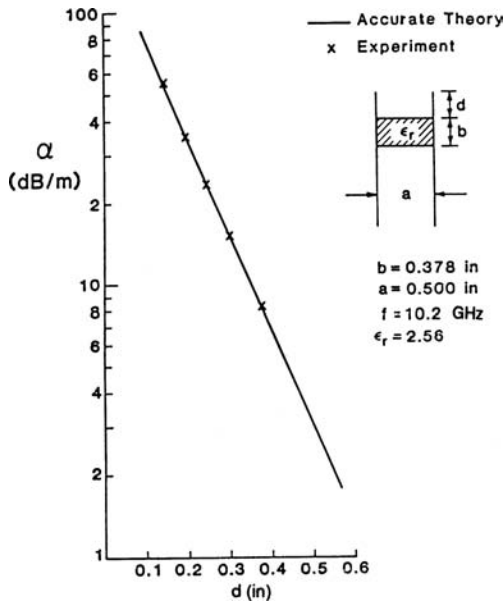


FIGURE 11-19 Comparison between measurement and theory for the leakage constant as a function of the length d of the foreshortened top for the NRD guide leaky-wave antenna shown in Figure 11-18 (after Han et al⁶⁹ © IEEE 1987)

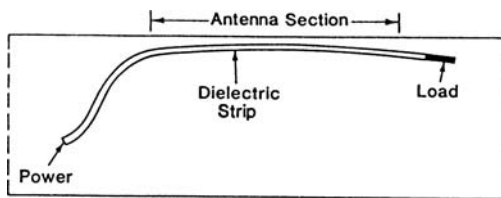


FIGURE 11-20 Side view of the foreshortened-top NRD guide leaky-wave antenna, showing how the value of α can be tapered very easily to reduce the sidelobes (after Sanchez and Oliner⁶⁷ © IEEE 1987)

Asymmetrical NRD Guide Antenna

The leakage mechanism for this NRD guide antenna is different from that for the antenna described just above, since this one is based on asymmetry, as shown in Figure 11-21. The structure shown on the left-hand side of Figure 11-18 is first bisected horizontally to provide radiation from one end only. Since the electric field is purely vertical in this midplane, the field structure is not altered by the bisection. An air gap is then introduced into the dielectric region, as shown, to produce asymmetry. As a result, a small amount of net horizontal electric field is created, which produces a mode in the parallel-plate air region akin to a TEM mode. This mode then propagates at an angle between the parallel plates until it reaches the open end and leaks away. It is necessary to maintain the parallel plates in the air region sufficiently

long that the vertical electric field component of the original mode has decayed to negligible values at the open end. Then the TEM-like mode, with its horizontal electric field, is the only field left, and the field polarization is then essentially *pure*. (The discontinuity at the open end does not introduce any cross-polarized field components.)

The asymmetry mechanism applied to the NRD guide in this way furnishes another leaky-wave antenna of simple configuration, but now with pure horizontal polarization. The air gap does not have to be large to produce a significant leakage rate. The geometry can thus be controlled easily to achieve a large range of values for α and, therefore, a large range of desired beamwidths. Furthermore, the air gap affects β only slightly, so that β and α can be controlled relatively independently. Of course, the guide must be operated in the fast-wave range to produce radiation.

The theory^{70,71} for this structure employs mode matching at the air-dielectric interfaces, and numerical results were obtained as a function of the geometric parameters. A detailed presentation of the theory and various numerical results appear in Chapter III of a comprehensive report.⁶³ No measurements have been taken to verify the theory.

There are many ways in which asymmetry can be introduced; this one was chosen for inclusion here because theoretical results are available for it. Measurements are available for a structure in which the asymmetry was produced by sloping the upper dielectric-air interface rather than introducing an air gap. Those measurements⁷² shown that radiation indeed occurs and that good patterns result, but parameter optimization has not yet been accomplished and no theory is available.

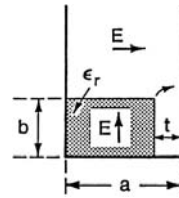


FIGURE 11-21 The asymmetrical NRD guide leaky-wave antenna produced by creating an asymmetrical air gap of thickness t and bisecting the NRD guide horizontally. The modifications introduced in the electric fields are also shown (after Oliner,⁶³ with Xu).

Stub-Loaded Rectangular Waveguide Antenna

The leaky-wave antenna whose cross section is shown in Figure 11-22 was derived from earlier work on groove guide, and its early name was the *offset-groove-guide antenna*.

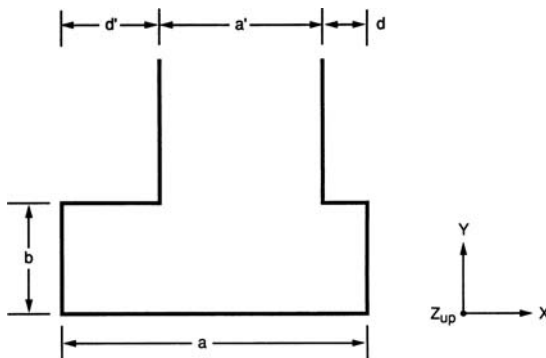


FIGURE 11-22 Cross section of the stub-loaded rectangular waveguide antenna, which has a remarkable set of important virtues. This antenna was earlier called the offset-groove-guide antenna (after Oliner,⁶³ with Lampariello and Shigesawa).

Under that name several presentations were made,⁷³⁻⁷⁵ and a full description was given in Chapters VII and VIII of Oliner.⁶³ Numerical results were derived for several different cross-sectional ratios, and some of the best performance was found for an aspect ratio corresponding to that for rectangular waveguide. The structure is also easy to feed from rectangular waveguide. For these reasons, the name was changed to the *stub-loaded rectangular waveguide antenna*, and three comprehensive companion papers⁷⁶⁻⁷⁸ were written with the new name.

Groove guide is a low-loss open waveguide for millimeter waves somewhat similar to the H guide or NRD guide; the dielectric central region of NRD guide is replaced by an air region of greater width. The fields are again strongest in the central region, and they decay exponentially vertically in the regions of narrower width above and below. Groove guide was popular some years ago, but it lost its glamour when it was realized that its higher-order modes would radiate, so that discontinuities in the guide would have some resistive content. This feature limited the range of applications for which groove guide was suitable, but it did not influence its application to leaky-wave antennas.

This antenna has a remarkable set of unusual virtues. One of the virtues is its versatility, which may be understood from the following considerations. When the stub is centered, the structure becomes nonradiating; alternatively, it may be viewed then as a slotted section cut in rectangular waveguide. For small off-center positions of the stub, the leakage rate will be small, yielding radiated beams of narrow width. When the offset is increased, α will increase, and the beamwidth will increase. We therefore have a relatively simple leaky-wave antenna, easily fed from a rectangular waveguide, that permits great versatility with respect to beamwidth by simply adjusting the location of the stub guide.

Even more important, it is found that the value of β changes very little as the stub is moved, while α varies over a very large range. This feature, namely, that β remains almost constant while α varies, makes it easy to taper the antenna aperture to control sidelobes. Thus it should be easy with this antenna to design beams with low sidelobe levels.

An additional advantage follows from the fact that the antenna is filled with only one medium, namely, air. As was shown in Section 11.2, the beamwidth then remains constant when the beam is scanned in elevation as you change the frequency.

The antenna was analyzed using a transverse equivalent network based on a new E -plane T -junction network.⁷⁹ This T -junction network is notable in that the expressions for the network elements are all in simple closed form and yet are very accurate. The resulting transverse equivalent network for the antenna is seen in Figure 11-23. Since the network elements are in closed form, the resonance relation for the complex propagation wave number is also in closed form, making calculations quick and easy.

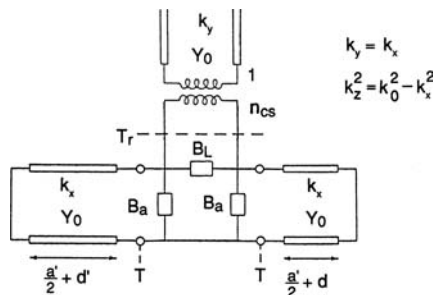


FIGURE 11-23 The transverse equivalent network for the antenna structure in Figure 11-22. Closed-form expressions for the network elements are given in the text.

The expressions for the elements of the network in Figure 11-23 may be written as

$$\frac{B_a}{Y_0} = -\frac{\pi}{16} \frac{a'}{b} \frac{k_x a'}{2} J_0^2\left(\frac{k_x a'}{2}\right) \quad (11-36)$$

where J_0 is the Bessel function of zero order,

$$\frac{B_L}{Y_0} + \frac{1}{2} \frac{B_a}{Y_0} = \frac{1}{n_c^2} \left(\frac{k_x b}{\pi}\right) \left[\ln\left(1.43 \frac{b}{a'}\right) + \frac{1}{2} \left(\frac{k_x b}{\pi}\right)^2 \right] \quad (11-37)$$

$$n_c = \frac{\sin(k_x a' / 2)}{k_x a' / 2} \quad (11-38)$$

$$n_{cs}^2 = n_c^2 (a' / b) \quad (11-39)$$

$$\begin{aligned} \frac{B_L}{Y_0} = & \frac{1}{n_c^2} \left(\frac{k_x b}{\pi}\right) \left[\ln\left(1.43 \frac{b}{a'}\right) + \frac{1}{2} \left(\frac{k_x b}{\pi}\right)^2 \right] \\ & + \frac{\pi}{32} \frac{a'}{b} \left(\frac{k_x a'}{2}\right) J_0^2\left(\frac{k_x a'}{2}\right) \end{aligned} \quad (11-40)$$

so that the dispersion relation for the transverse wave number k_x becomes

$$\frac{1}{n_{cs}^2} + j \frac{B_L}{Y_0} + j \frac{\left[\frac{B_a}{Y_0} - \cot k_x \left(\frac{a'}{2} + d\right) \right] \left[\frac{B_a}{Y_0} - \cot k_x \left(\frac{a'}{2} + d'\right) \right]}{2 \frac{B_a}{Y_0} - \left[\cot k_x \left(\frac{a'}{2} + d\right) + \cot k_x \left(\frac{a'}{2} + d'\right) \right]} = 0 \quad (11-41)$$

where the wave number k_x is related to k_z , the result that we seek, by

$$k_z = \beta - j\alpha = \sqrt{k_0^2 - k_x^2} \quad (11-42)$$

These expressions, and the transverse equivalent network in Figure 11-23, assume that the stub guide is infinite in length. In practice, of course, the stub length is finite, and it should only be long enough that the vertical electric field (represented in the stub guide by the below-cutoff TM_1 mode, viewed vertically) can decay to negligible values, permitting essentially pure horizontally polarized radiation. Usually, the stub length need only be about a half wavelength or less if the stub is narrow. The finite stub length can be readily taken into account, and a detailed treatment of its effects is reported in Oliner,⁶³ but it produces only a small change in the numerical values obtained from the procedure given earlier.

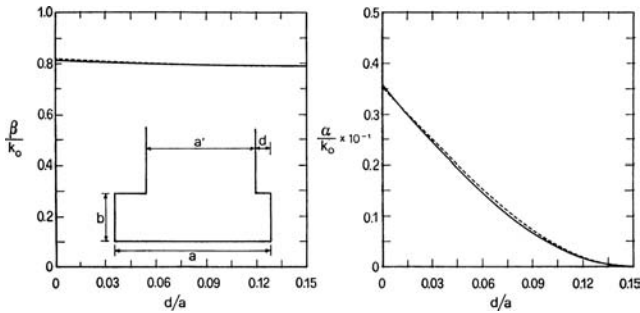


FIGURE 11-24 Comparison of numerical results for the stub-loaded rectangular waveguide antenna, obtained via two completely different theoretical methods. Additionally, we see that by varying the offset d , we can change α over a wide range of values while affecting β very little. Such independence makes it easy to taper the aperture distribution in order to control the sidelobes. $f = 28$ GHz, $a = 1.0$ cm, $a' = 0.7$ cm, $b = 0.3$ cm (after Oliner,⁶² with Shigesawa).

As an independent check on the accuracy of these expressions, the values of α and β were calculated using an entirely different theoretical approach, that of mode matching. As shown in Figure 11-24, where the dashed lines represent values obtained using the network and the solid lines represent those derived via the mode-matching procedure, the agreement is seen to be very good. We also can see from this figure that, as the stub is shifted laterally, β remains almost constant as α varies monotonically from zero to large values.

Measurements were also taken of the values of α and β and of the radiation patterns over the frequency range from 40 to 60 GHz.⁷⁸ A photograph of the cross section of the structure appears in Figure 11-25. The Japanese 500-yen coin (26.5 mm in diameter) is seen to dwarf the antenna cross section. The comparisons between the theoretical and measured values for both α and β are found to be very good.

The stub-loaded rectangular guide leaky-wave line source antenna is thus an attractive structure for millimeter wavelengths, since it is simple in configuration, easily fed, versatile in beamwidth, suitable for low-sidelobe-level designs, and capable of furnishing essentially pure horizontally polarized radiation. In addition, a simple and accurate theory is available for it that has been verified by an alternative, totally different computational approach, as well as by measurements.

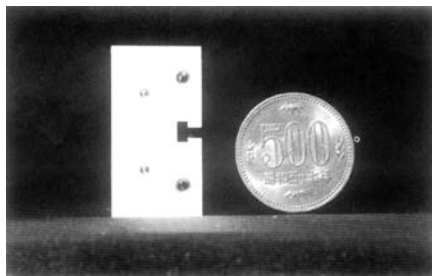


FIGURE 11-25 Photograph of the cross section of a stub-loaded rectangular waveguide antenna that was measured at millimeter wavelengths. A 500-yen Japanese coin (26.5 mm in diameter) is shown for size comparison (after Oliner,⁶³ with Shigesawa).

Printed-Circuit Version of Stub-Loaded Rectangular Waveguide Antenna

If the versatile leaky-wave antenna just described could be made in printed-circuit form, the fabrication process could make use of photolithography, and the taper design for sidelobe control could be handled automatically in the fabrication. That is, the location and width of the stub, and their variations along the antenna length in conformity with the sidelobe design requirements, could all be accomplished at the same time by either depositing the metal or etching some away to produce the gap. With this goal in mind, a printed-circuit version of the structure in Figure 11-22 took the shape shown in Figure 11-26.

The principal change is that the rectangular guide portion is now dielectric-filled so that the metal widths d' and d and the gap a' can be fabricated photolithographically. Also, the vertical baffle portions are now moved to the sides of the guide, as extensions of the guide side walls.

The basic physics of the leakage is the same as that for the antenna in Figure 11-22. The rectangular guide, now dielectric-filled, is fed from one end with the electric field vertical inside the guide. This vertical E field excites the asymmetrically located longitudinal slot such that the resultant field in the slot has both vertical and horizontal components. These field components in turn excite the TM_1 mode (viewed vertically) and the TEM mode (at an angle) respectively, in the parallel-plate air-filled stub region. The separation a between the stub walls is less than a half wavelength, so that the TM_1 mode is below cutoff; the TEM mode, being above cutoff, propagates at an angle to the upper end and radiates. The stub dimension c must then be sufficiently long that the field of the TM_1 mode can

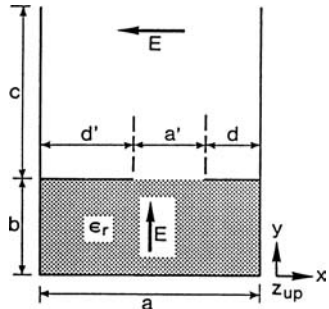


FIGURE 11-26 Printed-circuit version of the stub-loaded rectangular waveguide antenna shown in Figure 11-22. This version has the advantage that the tapering to control sidelobes can be effected photolithographically all at one time (after Oliner,⁶³ with Lampariello and Frezza).

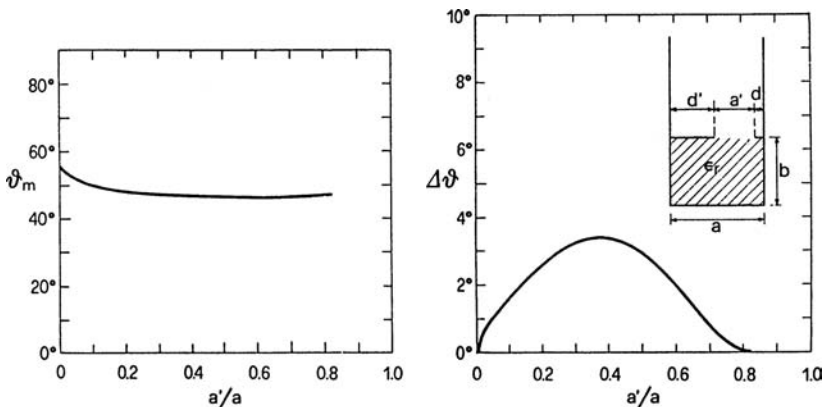


FIGURE 11-27 Variations of the beam angle θ_m and the beamwidth $\Delta\theta$ as a function of the relative slit width for the printed-circuit structure in Figure 11-26. It is seen that as a'/a varies from about 0.4 to about 0.8, the beamwidth changes over a wide range while the beam angle remains almost constant, showing that a' is a good parameter to vary to reduce the sidelobes. $f = 50$ GHz, $\epsilon_r = 2.56$, $a = 2.25$ mm, $b = 1.59$ mm, $d = 0.20$ mm (after Oliner,⁶³ with Lampariello).

decay to negligible values. As a result, the radiated power is polarized horizontally, with negligible cross-polarization. It would be nice to eliminate the vertical stub walls, but they are necessary for eliminating the vertical electric field component.

The transverse equivalent network for this new antenna structure is slightly more complicated than the one in Figure 11-23, and the expressions for the network elements must be modified appropriately to take the dielectric medium into account. The detailed derivation of the modified network and the expressions for the network parameters appears in Chapter X of Oliner.⁶³

Again, α can be varied by changing the slot location d . However, it was found that a' is also a good parameter to change for this purpose. The variations in beam-elevation angle θ_m and beamwidth $\Delta\theta$ with relative slot width a'/a are presented in Figure 11-27. You can see that as a'/a varies from about 0.4 to about 0.8, the beamwidth changes from about 4° to nearly zero but that the beam location moves hardly at all over that range. The slot width a' therefore represents a good parameter to vary to control the sidelobes.

This antenna structure has not yet been examined experimentally, but it appears to be a very promising one.

An interesting variation of the structures shown in Figures 11-22 and 11-26 has been developed and analyzed.⁸⁰ It is based on a ridge waveguide rather than a rectangular waveguide, and the structures can be either air-filled or dielectric-filled. In the structures based on rectangular waveguide, the asymmetry was achieved by placing the stub guide, or locating the longitudinal slot (or slit), off-center on the top surface. Here the top surface is symmetrical, and the asymmetry is created by having unequal stub heights on each side under the main guide portion.

By a clever arrangement of the geometry, the coauthors, Frezza, Guglielmi, and Lampariello,⁸⁰ were able to vary the values of α and β essentially independently by adjusting combinations of the vertical stubs. Let us define two geometric parameters: (1) the relative average arm length b_m/a , where $b_m = (b_l + b_r)/2$, and b_l and b_r are the left and right stub arm lengths; (2) the relative unbalance $\Delta b/b_m$, where $\Delta b = (b_l - b_r)/2$. It then turns out that by changing b_m/a , you can adjust the value of β/k_0 without altering α/k_0 much and that by changing $\Delta b/b_m$, you can vary α/k_0 over a large range without affecting β/k_0 much. In this way, by changing only the stub lengths, you can exercise essentially independent control over β and α . The taper design for controlling the sidelobe level would therefore involve only the relative unbalance, $\Delta b/b_m$.

The air-filled ridge waveguide structure may therefore have a constructional advantage regarding sidelobe control over the simpler structure shown in Figure 11-22. However, the dielectric-filled ridge waveguide antenna lacks the important advantage offered by the printed-circuit version in Figure 11-26, for which you can make a mask that permits the whole upper surface, including the taper for sidelobe control, to be realized photolithographically.

11.8 ARRAYS THAT SCAN IN TWO DIMENSIONS

All the leaky-wave antennas discussed so far in this chapter are line-source antennas that produce conical beams. They can scan in elevation (or azimuth if they are rotated, of course) as the frequency is varied, with their beamwidths narrow in elevation but wide in the cross-plane. General comments along these lines were made earlier, and it was pointed out that the beamwidth in the cross-plane could be narrowed by using a horn (see Figure 11-12) or by arranging a parallel set of these line sources in an array. Such arrays, however, do not scan in the cross-plane.

This section describes some of the most recent developments involving leaky-wave antennas, where an array of leaky-wave line sources is so arranged that it can scan in both elevation and azimuth. Such an array is really a linear phased array of leaky-wave

line-source antennas, where scanning in elevation is obtained in the usual leaky-wave fashion by varying the frequency, and scanning in the cross-plane (or azimuth) is achieved by varying the phase difference between the successive parallel leaky-wave line sources. The approach itself is not new, but the specific structures are. As will be seen, these arrays have some unusual virtues: essentially no cross polarization, no grating lobes, no blind spots within the scan volume, and simplicity of structure. In addition, they should cost less than phased arrays because phase shifters are needed in only one dimension. These different array structures will be discussed, but first we present the principle of operation (the basic architecture for such arrays) and some comments about the analytical approach, which takes into account all mutual coupling effects.

Principle of Operation

This class of scanning arrays achieves scanning in two dimensions by creating a one-dimensional phased array of leaky-wave line-source antennas. The individual line sources are fed from one end and are scanned in elevation by varying the frequency (or by electronic means if and when available). Scanning in the cross-plane, and therefore in azimuth, is produced by phase shifters arranged in the feed structure of the one-dimensional array of line sources. This frequency-phase scan architecture is illustrated schematically in Figure 11-28.

The radiation will therefore occur in pencil-beam form and will scan in both elevation and azimuth in a conical-scan manner. The spacing between the line sources is chosen such that no grating lobes occur, and accurate analyses show that no blind spots appear anywhere. The leaky-wave line-source antennas employed in the three examples of arrays described here are ones we have already discussed earlier or are modifications of them. The advantage of negligible cross polarization at all angles follows from the fact that the individual line-source antennas possess that feature and that the array arrangement does not introduce any cross-polarized components.

In principle, a large variety of different leaky-wave line sources can be used in this array architecture. In fact, however, you must be very selective here, because the line sources must be integrated into the overall geometry if the resulting antenna is to remain simple in configuration. The list is further limited to those structures amenable to analysis because the design requires a theoretical basis. If the radiating portion of a suitable structure in the class can be fabricated by photolithographic means, using a mask, the costs can also be kept down, and the method is amenable to mass-production techniques. Two of the three arrays to be described fall into this category.

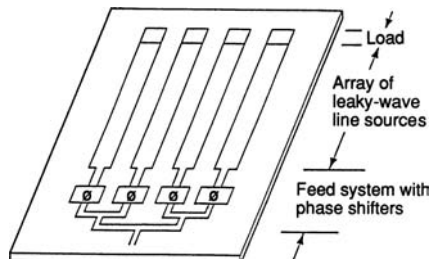


FIGURE 11-28 Schematic of the recent approach to simpler 2D scanning, involving a linear phased array of leaky-wave line sources, with frequency scanning in elevation and phase scanning in azimuth (after Oliner,⁶² with Peng).

Analytical Approach

The arrays to be described have been analyzed accurately by a *unit-cell approach* that takes into account all mutual coupling effects. Each unit cell incorporates an individual line-source antenna, but in the presence of all the others. These individual line-source antennas were analyzed using a transverse equivalent network in which the radiating open end was representative of the environment of the *single* lone line source. In the *array* of such line sources, the radiating environment is, of course, quite different, and it will change as the array is scanned in azimuth.

The treatment of the periodic external environment by a unit-cell approach automatically accounts for all mutual coupling effects and provides information on all the effects of scan. The radiating termination on the unit cell modifies the transverse equivalent network, and the resonances of this transverse network yield the properties of the leaky wave guided along the line sources. A key new feature of the array analysis is therefore the determination of the *active admittance* of the unit cell in the 2D environment as a function of scan angle. This active admittance is the input admittance to the external radiating region, and it is appended to the remainder of the transverse equivalent network, the latter being different for each of the arrays to be discussed.

Array of Asymmetrical NRD Guide Line Sources

A cross-sectional view of the linear phased array of asymmetrical NRD guide leaky-wave line sources appears in Figure 11-29. By comparison with the individual line-source antenna in Figure 11-21, it is clear that the array consists of a number of these line sources placed directly next to each other. The line sources provide the elevation pattern, modified by their presence in the array, and the geometry in the plane shown specifies the cross-plane behavior, with the angle of scan determined by the phase shift imposed between successive line sources.

The unit cell for this array is shown in Figure 11-30, where the phase-shift-wall properties depend on the scan angle in the cross-plane. Thus the discontinuity at the end of the stub of length c , and therefore the values of β and α , depend on that scan angle. The analysis proceeds, therefore, by knowing the imposed phase shift between line sources, and, from it and the unit-cell network, by finding β and α in the axial direction. Everything else follows directly from this information.

If the values of β and α did not change with phase shift, the scan would be exactly conical. However, it is found that these values change only a little, so that the deviation from conical scan is small.

We next consider whether blind spots are present. *Blind spots* refer to angles at which the array cannot radiate or receive any power; if a blind spot occurred at some angle, therefore,

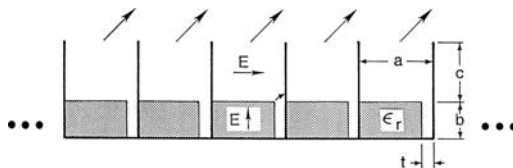


FIGURE 11-29 Cross section of linear phased array of asymmetrical NRD guide leaky-wave line sources. The individual line source appears in Figure 11-30 (after Oliner,⁶³ with Xu).

the value of α would rapidly go to zero at that angle of scan. To check for blind spots, we would then look for any sharp dips in the curves of α/k_0 as a function of scan angle. No such dips were ever found. Typical data of this type exhibit fairly flat behavior for α/k_0 until the curves drop quickly to zero as they reach the end of the conical scan range, where the beam hits the ground.

The theoretical analysis and many numerical results appear in Chapter IV of Oliner,⁶³ and a presentation was made.⁸¹

Array of Printed-Circuit Uniform Line Sources

If the leaky-wave line sources in the array were in printed-circuit form, the fabrication process could make use of photolithography, and the taper design for sidelobe control could be handled automatically in the fabrication. The array structure shown in Figure 11-31 fits into this category. It may be seen that the line-source elements in this array are exactly the printed-circuit version of the stub-loaded rectangular waveguide antenna shown in Figure 11-26. The leakage rate, and therefore the beamwidth, can again be controlled by varying the width or location of the gap within each element. And again, the value of β remains almost constant as the gap width or location is varied, so that a taper design for sidelobe control is easy to implement.

The metal fins that project vertically, which may alternatively be called *baffles* or *stubs*, serve two purposes. The first purpose is to ensure *essentially pure horizontally polarized radiation*, with negligible cross polarization. As explained in the discussion associated with Figure 11-27, the fins form a stub guide that is below cutoff for the vertical electric field, thereby permitting only the horizontal electric field to radiate.

The second purpose of the stubs is to *eliminate blind spots*. With the stubs present, we have never found any, and it is known that arrays of this sort with dielectric layers often exhibit them when there are no stubs. A careful examination shows that the stub length should be roughly a half wavelength, which means that the projection is actually rather small, particularly at millimeter wavelengths.

The transverse equivalent network for the unit cell representing the array in Figure 11-31 is very similar to the one for the isolated line-source antenna and differs from it only in the terminating admittance placed on the end of the stub line. This difference is important,

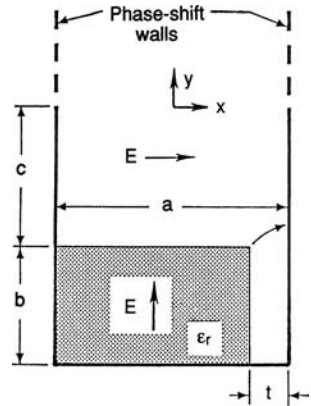


FIGURE 11-30 Unit cell of the linear array of NRD guide line-source antennas shown in Figure 11-29. The phase-shift walls change with scan angle, and their use leads to an analytical approach that takes all mutual coupling effects into account automatically (after Oliner,⁶³ with Xu).

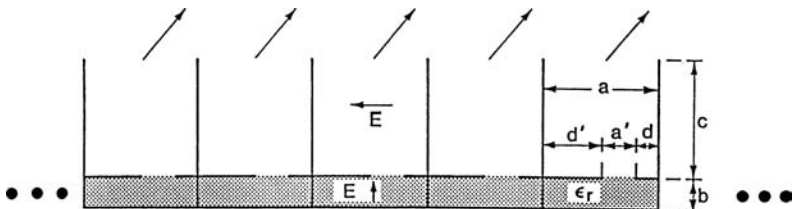


FIGURE 11-31 Cross section of the linear phased array of printed-circuit *uniform* leaky-wave line sources. The individual line-source antenna appears in Figure 11-26 (after Lampariello and Oliner⁸²).

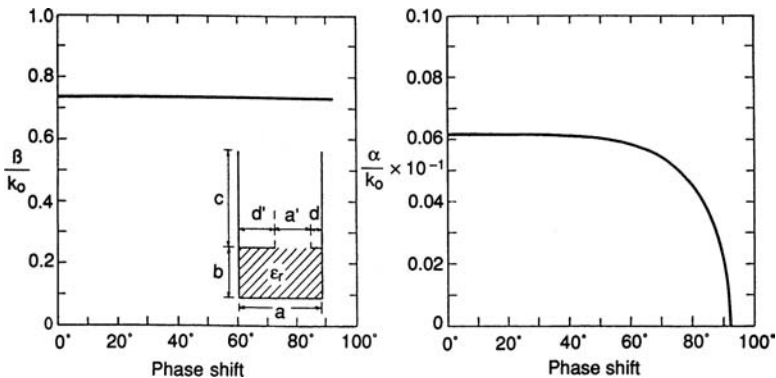


FIGURE 11-32 Variations of the normalized phase constant and leakage constant with phase shift between adjacent line sources (cross-plane scan) for the array appearing in Figure 11-31, showing conical scan and no blind spots. $f = 50$ GHz, $\epsilon_r = 2.56$, $a = 2.25$ mm, $a' = 1.00$ mm, $b = 1.59$ mm, $c = 6.00$ mm, $d = 0.25$ mm (after Lampariello and Oliner⁸²)

however, in that the terminating admittance is a function of scan angle and takes into account the mutual coupling effects of all the neighbors in the array.

As in the case of the array of asymmetrical NRD guide line sources, the value of β/k_0 changes very little with the phase shift between successive line sources, so that as a result, there is little deviation from conical scan. Also, the curve of α/k_0 as the phase shift is changed does not show any sharp dips but remains fairly flat until it drops as the beam approaches the ground at the end of the scan range. Thus no blind spots have ever been observed. Calculated curves for one set of geometric parameters that demonstrate this behavior are given in Figure 11-32.

A summary of the results obtained for this array appears in Lampariello and Oliner,⁸² further results and the details of the theoretical derivation are contained in Chapter X of Oliner.⁶³

Array of Printed-Circuit Periodic Line Sources

The purpose of employing arrays with *periodically* loaded leaky-wave line sources, instead of *uniform* ones like the two discussed earlier, is to *increase the scan-angle coverage*, as will be explained. An example of such an array is given in Figure 11-33. This array can be fed by using a series of dielectric-filled rectangular guides placed on their sides, so that the periodic elements are excited with a horizontal electric field and the array will radiate with pure horizontal polarization.

The individual leaky-wave line sources in this array may at first look like the periodic metal-strip antenna shown in Figure 11-2 and discussed in Section 11.6, but they are not the same because these are excited by a horizontal electric field, parallel to the strips, whereas the antenna in Figure 11-2 is excited by an electric field that is primarily vertical.

The structure in Figure 11-33 is shown with stub guides, or baffles, present.

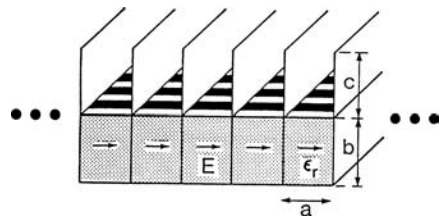


FIGURE 11-33 Linear phased array of printed-circuit *periodic* leaky-wave line sources, which is capable of a greater scan range than is possible with the arrays containing uniform line sources (after Oliner,⁶³ with Guglielmi)

For this array, the stubs are not necessary to ensure pure horizontally polarized radiation because no component of vertical field is excited. On the other hand, they also ensure that no blind spots will occur. It is possible that no blind spots will be present when the stubs are removed, but we do not know the answer. Certainly, the structure is simpler and easier to fabricate without the stubs. Either way, the printed-circuit nature of the strips on the dielectric surface permits the detailed metal circuit, including the tapering for sidelobe control, to be deposited photolithographically on the dielectric surface.

As discussed elsewhere in this chapter, leaky-wave antennas with uniform apertures can radiate only into the forward quadrant, and they cannot scan too close to broadside or end fire. In the present periodically loaded array, the dominant mode is chosen to be slow so that it is purely bound; the period is then selected relative to the wavelength so that only the $n = -1$ space harmonic becomes radiating. As a result, the array provides greater scan coverage, since the $n = -1$ beam can scan over the complete backward quadrant and part or all of the forward quadrant, depending on the parameters (except for a narrow angular region around broadside).

In the design of this array, it should be remembered that these small slits radiate very weakly. The discussion with respect to both theory and numerical results in Chapter XI of Oliner⁶³ is restricted to such narrower slits and, therefore, narrower beams. A presentation on this antenna was given.⁸³

The three arrays described earlier are examples of how a linear phased array of leaky-wave line sources can be arranged to provide 2D scanning of a pencil beam over a restricted sector of space. In addition, these three arrays possess certain special virtues. They provide essentially pure horizontally polarized radiation, and they exhibit no blind spots or grating lobes.

11.9 NARROW-BEAM ANTENNAS BASED ON A PARTIALLY REFLECTIVE SURFACE

Overview

This class of leaky-wave antennas is based on planar technology, and it offers a simple way to obtain high directivity with a small source. This category is a relatively new one, and the number of antennas that fit into it has grown slowly over time. These antennas bear some resemblance to the classical Fabry-Perot structure in the optics field, but very few actual antennas have developed from it. In fact, as we shall show, in one important case very similar structures have emerged from very different approaches.

The antenna structures in this section are different from those discussed in the rest of this chapter, but they are also leaky-wave antennas, and their performance obeys the fundamental relations described in the earlier sections of this chapter. These antennas are planar in nature, and most of them consist of a metal ground plane with a dielectric layer on it (which may be air) that is covered with a *partially reflective surface or screen* (PRS) on top of the dielectric layer. The partially reflective (or partially transparent) screen can take various forms, such as a stack of one or more dielectric layers, or a metal screen consisting of a periodic array of slots or patches, or an array of parallel wires. Several examples of such structures are shown in Figure 11-34 (where an electric dipole source is shown). The structure is typically excited by a simple source inside the dielectric layer, which may, for example, be a horizontal electric dipole in the middle of the layer or a magnetic dipole on the ground plane. Such sources may be realized in practice by using printed dipoles inside the layer, microstrip patches on the ground plane, slots in

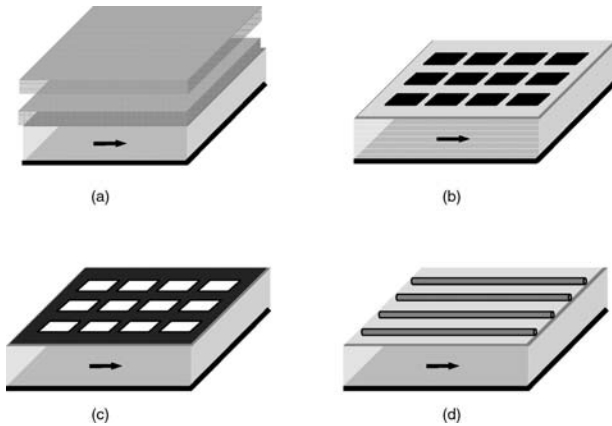


FIGURE 11-34 Examples of PRS-based leaky-wave antennas: (a) The PRS consists of a stack of dielectric layers. (b) The PRS consists of a periodic array of metal patches. (c) The PRS consists of a periodic array of apertures in a metal plate. (d) The PRS consists of a periodic array of closely spaced wires. In each case a dipole source excitation is shown inside the dielectric layer.

the ground plane, or waveguide-fed apertures in the ground plane. Since the fundamental nature of the beamforming is through leaky-wave radiation, the source merely acts as a launcher for the leaky waves.

For a horizontal dipole source, the radiation produced may be a narrow pencil beam pointing at broadside ($\theta_0 = 0$), or a conical beam pointing at any desired scan angle $\theta_0 > 0$. As discussed later, the thickness of the dielectric layer controls the scan angle. For a conical beam, the pattern is usually fairly omnidirectional (azimuthally independent) for small angles θ_0 , but the E - and H -plane beamwidths typically become more different as the scan angle increases. A vertical dipole source can only produce a conical beam (which is omnidirectional) at a scan angle $\theta_0 > 0$. An illustration of these two types of patterns (broadside and conical) is shown in Figure 11-35.

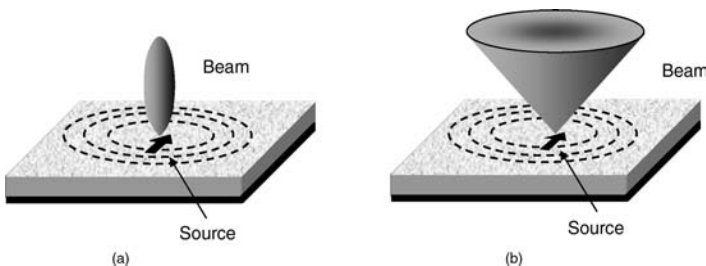


FIGURE 11-35 An illustration of the beam types that can be realized by using a PRS leaky-wave antenna excited by a horizontal dipole source: (a) a pencil beam at broadside and (b) a conical beam pointing at an angle greater than zero. The dipole source launches a radially propagating cylindrical leaky wave, the phase fronts of which are shown by the dashed lines.

Basic Principles of Operation

One of the main differences between this type of structure and the ones considered previously is that the leaky wave on this structure is a 2D cylindrical wave, which propagates outward radially from the source along the interface.⁸⁴ The leaky wave then furnishes a large aperture that in turn produces the narrow radiation beam. As is true for all leaky-wave antennas, the narrow beam angle and the beamwidth are frequency sensitive. A vertical electric or magnetic dipole source launches only a TM_z or TE_z leaky wave respectively, which has no ϕ variation. This results in an omnidirectional conical beam.⁸⁴ A horizontal electric or magnetic dipole source launches a pair of leaky waves, one TM_z and one TE_z . The TM_z leaky wave determines the E -plane pattern, while the TE_z leaky wave determines the H -plane pattern.⁸⁴ For a broadside beam these two leaky waves have very nearly the same phase and attenuation constants, and hence an omnidirectional pencil beam is created. Interestingly, this is true even if the PRS is not similar in the E - and H -plane directions. For example, the PRS may consist of a periodic array of slots in a metal plate, with the slots being long in the x direction and narrow in the y direction, having very different periodicities in the two directions. As the scan angle increases, the pencil beam turns into a conical beam, similar to how the petals on a flower unfold. As the scan angle increases (by increasing the frequency or the layer thickness) the wavenumbers of the two leaky waves typically begin to differ, and this explains why the beamwidths often become different in the principal planes. (The exact nature of the beamwidth variation with scan angle depends on the particular type of PRS.)

The PRS is used to create a leaky parallel-plate waveguide region, and the leaky waves are leaky (radiating) versions of the parallel-plate waveguide modes that would be excited by the source in an ideal parallel-plate waveguide, which results if the PRS is replaced by a perfectly conducting metal plate. This point of view allows for a simple design formula for the thickness of the dielectric layer in order to obtain a beam at a desired angle θ_0 (either a broadside or a conical beam). The parallel-plate waveguide modes are described by $n = 1$, meaning that there is one half-wavelength variation vertically inside the parallel-plate waveguide. (Although larger values of n could be used, this would result in a design that has a thicker dielectric layer.) The radial wavenumber of the TM_z and TE_z parallel-plate waveguide modes for an ideal waveguide would be

$$k_\rho = \beta = \sqrt{k_1^2 - \left(\frac{\pi}{h}\right)^2} \quad (11-43)$$

where h is the thickness of the dielectric layer and k_1 is the wavenumber of the layer, which may also be expressed as $k_1 = k_0 n_1$, where n_1 is the refractive index of the layer. Using the simple relation $\beta = k_0 \sin \theta_0$ that is valid for any leaky wave, we obtain the result

$$h = \frac{\lambda_0 / 2}{\sqrt{n_1^2 - \sin^2 \theta_0}} \quad (11-44)$$

The location of the source usually has little effect on the pattern shape, since this is dictated by the leaky-wave phase and attenuation constants. The phase constant is primarily determined by the thickness of the dielectric layer (see Eq. 11-43), while the attenuation constant is primarily determined by the properties of the PRS. However, the power density at the peak of the beam will be maximum when a horizontal electric dipole source is placed in the middle of the dielectric layer, or a horizontal magnetic dipole is placed on the ground plane.

A vertical electric dipole source maximizes the peak power density when it is placed on the ground plane, while a vertical magnetic dipole source does so when it is placed in the middle of the layer. Changing the peak power density, and hence the overall power radiated by the source, directly affects the input resistance of the source.

The PRS may be either uniform (e.g., one or more dielectric layers) or periodic in one or two dimensions (e.g., an array of closely spaced wires, or a 2D array of slots in a metal plate). However, it is important to note that for the periodic PRS structures, the radiation still occurs via the fundamental parallel-plate waveguide modes, and not from a space harmonic of these modes. That is, the PRS acts as a quasi-uniform reflective surface, where the reflection coefficient of the fundamental parallel-plate wave determines the characteristics of the antenna. The physical principle of operation is thus as a quasi-uniform leaky-wave antenna, and not a periodic leaky-wave antenna that radiates from a space harmonic.

To our knowledge, the first use of a quasi-uniform PRS to improve the antenna gain was described by von Trentini⁸⁵ in 1956. A sketch of that structure was presented by the author and is repeated here as Figure 11-36, in which he placed a source at P on a ground plane and then located a PRS parallel to the ground plane a distance l in front of it. He views the performance in terms of multiple reflections between the ground plane and the PRS, and then derives an expression for the resonance condition that yields maximum radiated power at broadside. In this derivation, he assumes that the PRS consists of an array of closely spaced parallel conducting wires oriented parallel to the electric field. With this model, he calculated the radiation patterns for several sets of dimensions. He also built and measured several antennas based on this PRS, and on others, including one consisting of an array of closely spaced circular holes in a metal plate. The source employed was a rectangular waveguide aperture, and measurements were made at a wavelength of 3.2 cm. The measurements and calculations for the main lobe of the radiation pattern agreed well with each other.

An improved version of the von Trentini antenna has been developed recently by Feresidis and Vardaxoglou.⁸⁶ These authors followed the von Trentini ray theory analysis, assuming the structure to have infinite extent, and employed a feed consisting of a rectangular waveguide built into the ground plane. The new contribution, and the “optimization” to which they refer, applies to their PRS structure. They note that the antenna would have greater bandwidth if the phase of its PRS were to linearly increase with frequency, in effect compensating somewhat for the path length that the rays must traverse (or equivalently for the change in the electrical thickness of the dielectric layer—see Eq. 11-44). They therefore investigated PRSs loaded with several different elements, such as crossed dipoles, patches, rings, and square loops. They found that dipoles, or square or circular patches (or their complementary structures), particularly with close packing of the elements in the array, produced less of a variation of the beam with frequency. This slower variation was not found for crossed dipoles, square loops, and rings, even for close packing of these elements. They therefore chose to use arrays of closely spaced dipoles in their PRS.

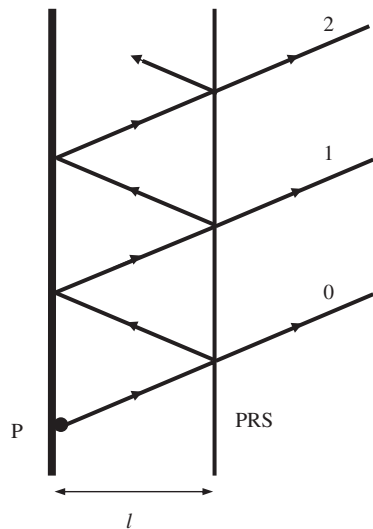


FIGURE 11-36 A ray explanation for the PRS-based leaky-wave antenna, as originally introduced by von Trentini (after von Trentini⁸⁵ © IEEE 1990)

Dielectric-Layer PRS Structures

Printed-circuit antennas, such as microstrip antennas, have many advantages for conformal applications, but they have the major disadvantage of low gain. During 1984 and 1985, Alexopoulos and Jackson published a pair of papers^{87,88} showing that the gain could be enhanced significantly by placing a dielectric superstrate, or cover layer, over the original dielectric layer. By choosing the layer thicknesses and the dielectric constant values appropriately, a large gain can be obtained for radiation at any desired angle, in the form of a pencil beam at broadside or a conical beam pointing at an angle θ_0 . The first of these two papers showed that a properly designed two-layer (substrate/superstrate) structure could produce such beams. The later paper⁸⁸ by Jackson and Alexopoulos examined in greater detail the quantitative relationships between the radiation properties of the two-layer dielectric structure and the parameter values (dimensions and permittivities) of the structure. Neither paper recognized the gain enhancement effect as due to leaky modes, however. The substrate-superstrate geometry of the structure, identifying the various parameters, is shown in Figure 11-37. A horizontal electric dipole source is shown located in the substrate and is parallel to the ground plane and the dielectric interfaces.

The superstrate layer will act as an optimum reflecting surface when the thickness is chosen so that it is an odd multiple of one-quarter of a wavelength in the vertical direction. This corresponds to the condition

$$t = \frac{(2m-1)\lambda_0}{4\sqrt{n_2^2 - \sin^2 \theta_0}} \quad (11-45)$$

where m is a positive integer and n_2 is the refractive index of the superstrate. Using $m = 1$ gives the thinnest superstrate.

An important next step was to recognize that the two-layer structure can support *leaky waves*, and that the structure can be analyzed as a *leaky-wave antenna*. A detailed analysis appeared in 1988 in a paper by Jackson and Oliner,⁸⁹ in which it was established that the directive beams obtained in the two-layer dielectric substrate/superstrate structure are due to the excitation of leaky modes by the horizontal dipole within the substrate layer. The leaky modes travel radially outward from the source and are supported by the two-layer structure. In particular, it was verified that a TM_z leaky mode determines the E -plane pattern, while a TE_z leaky mode determines the H -plane pattern. The trajectory of these two leaky modes in the steepest-descent plane illuminated very clearly how the radiated beam, as a combination of these two leaky modes, behaves as a function of frequency. The leaky-mode approach, which automatically takes into account how the leaky modes decay as they travel away from the dipole source, also tells us how large the ground plane and layer structure must be to act as if it is effectively infinitely wide. The leaky-wave approach is the only one that furnishes such information.

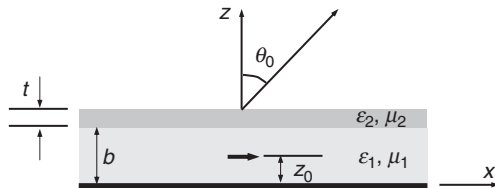


FIGURE 11-37 The two-layer substrate/superstrate configuration (after Alexopoulos and Jackson⁸⁷ © IEEE 1984)

The directivity increases as the permittivity of the superstrate layer increases relative to that of the substrate layer, since the superstrate PRS then acts as a more reflective surface. Another way to increase the reflection from the PRS was proposed by Yang and Alexopoulos⁹⁰ in 1987, in which the single superstrate is replaced by a *periodic array of such superstrates*, as shown in Figure 11-38. In this structure the PRS consists of a stack of multiple superstrate layers, where the high-permittivity superstrate layers with parameters (ϵ_2, μ_2) are separated by low-permittivity spacer layers with parameters (ϵ_1, μ_1). The high-permittivity layers have thickness t chosen from Eq. 11-45, while the low-permittivity spacer layers have one-half the thickness of the bottom dielectric layer. By using the multiple superstrates, the bottom layer is further isolated from the free-space region above, and the radiated beam becomes even narrower. The directivity increases geometrically with the number of superstrate layers, and thus very directive beams may be obtained using modest values of superstrate permittivity, provided several superstrate layers are used. A leaky-wave explanation for the multiple-layer dielectric structure proposed by Yang and Alexopoulos was presented in 1993 by Jackson et al.⁹¹ Formulas were derived for the leakage constant as a function of the layer parameters and the number of superstrate layers.

A fundamentally different approach to analyzing the multiple-superstrate structure was published in 1999 by Thevenot et al.⁹² The structure is the same as the one seen in Figure 11-38, but for only two superstrate layers. The source is also a patch antenna rather than a dipole, and the low-permittivity region is air. But the main point here is that the *approach* and its associated *terminology* are novel and completely different. In the adopted approach, a periodic array of low/high permittivity superstrate layers, which serves as the cover for the bottom dielectric layer, operates as a *photonic-band-gap* (PBG) structure, meaning that it is in a stopband. According to the authors, the design procedure is to “open an extra mode inside the frequency gap by inserting a defect into the dielectric period. The directive antenna described in this section uses the defect mode of a photonic crystal to achieve electromagnetic radiation. The PBG material is used as a cover to enhance the gain of a usual patch antenna.” Theoretical calculations were made of the radiation patterns of the patch antenna with and without the presence of the PBG cover. With the cover present, the directivity increased from 8 to 20 dB. Calculations were also made using an FDTD code, and then compared with measurements taken on an antenna structure built using a cover consisting of two layers of alumina rods. The comparison was quite good.

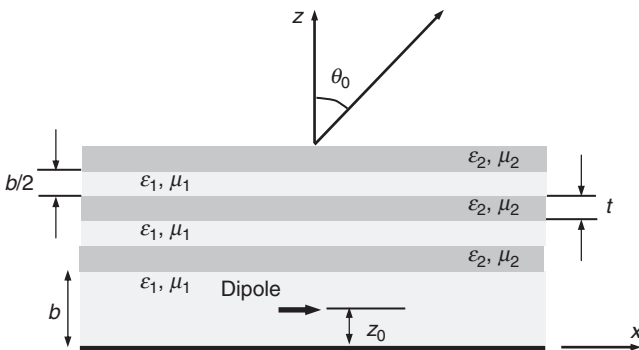


FIGURE 11-38 The multiple superstrate configuration, in which an alternating stack of low- and high-permittivity layers is used to form the PRS (after Yang and Alexopoulos⁹⁰ © IEEE 1987)

Periodic PRS Structures

As noted earlier, the PRS may consist of a periodic structure that is periodic in one direction, such as an array of closely spaced wires (or more generally, a stack of wire layers). The PRS may also consist of a 2D periodic array of elements. As also mentioned, the original work of von Trentini consisted of such PRSs. Here we examine the radiation properties for two particular PRSs, one consisting of a periodic array of metal patches,⁹³ and the other the complement, namely a periodic array of rectangular apertures or slots in a metal plate.⁹⁴

The metal-patch PRS structure is shown in Figure 11-34*b*, while the slot-PRS structure is shown in Figure 11-34*c*. Both PRSs may be used to produce a broadside pencil beam when air is used as the dielectric. Figure 11-39 shows a typical set of patterns obtained with the patch-PRS structure, showing the H -plane pattern for a broadside beam and a conical beam at $\theta_0 = 45^\circ$. The E -plane patterns, not shown, are similar. For a conical beam pointing at an angle $\theta_0 > 0$, the beam angle θ_0 is limited to about 45° . This is because the thickness h of the air layer increases without limit as the angle θ_0 increases toward 90° , as seen from Eq. 11-44. As the thickness increases beyond one wavelength, the next higher-order set of TM_z and TE_z parallel-plate modes ($n = 2$) begin to propagate, and radiation from this set of leaky modes results in an undesirable secondary beam.

The limitation on the beam angle due to the higher-order parallel-plate modes may be overcome by using a dielectric layer with $\epsilon_r > 4/3$, which allows for the beam angle θ_0 to approach 90° before radiation from a secondary beam occurs.⁹³ However, when using a dielectric layer with the patch-PRS, it is observed that undesirable secondary beams arise from another source, namely from the -1 space harmonic of the (perturbed) TM_0 surface wave that is supported by the grounded dielectric layer (which is perturbed by the metal patches).⁹³ This problem is avoided when using the slot-PRS structure since the guiding structure is now a perturbed parallel-plate waveguide (perturbed by the apertures) rather than a perturbed grounded slab. A beam approaching 90° may be realized by using a slot-PRS structure with a dielectric layer.⁹⁴ However, as the beam angle increases, the E - and H -plane patterns become increasingly different, with the E -plane pattern broadening and the H -plane pattern narrowing.

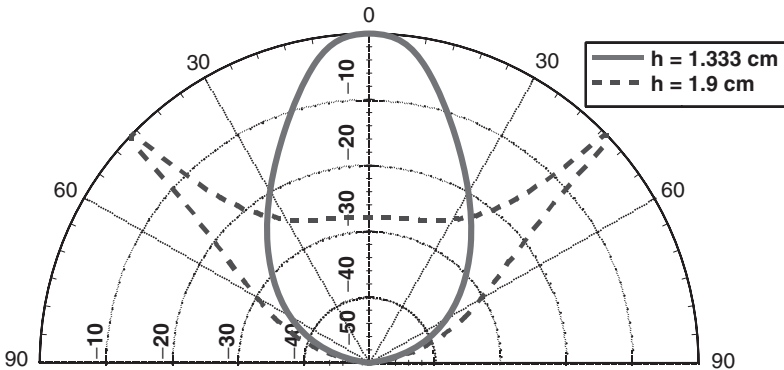


FIGURE 11-39 H -plane radiation patterns (in dB) at 12 GHz for a broadside beam and a 45° conical beam, for the patch-PRS structure of Figure 11-34*b*. The patches have lengths (x -dimension) of 1.25 cm and widths (y -dimension) of 0.01 cm, with periodicities of 1.35 cm in the x direction and 0.3 cm in the y direction. The thickness of the dielectric layer is $h = 1.33$ cm ($0.534\lambda_0$) for the broadside beam and $h = 1.90$ cm ($0.761\lambda_0$) for the conical beam. The structure is excited by an x -directed infinitesimal horizontal electric dipole in the middle of the layer (after Zhao et al⁹³ © IEEE 2005).

Very recent work has shown that it is possible to overcome the problem of different beam behaviors in the E - and H -planes by the use of a wire PRS (an array of closely spaced conducting wires) together with a dielectric layer that is air or a low-permittivity material.⁹⁵ Evidently, this is because the characteristics of the wire PRS change with the angle θ_0 in a manner so as to compensate for the natural change in the different principal planes as the beam angle changes (see the discussion in the next subsection).

General Design Formulas

The far-field pattern of a PRS leaky-wave antenna structure may be calculated by reciprocity, in which the far field is determined by illuminating the structure with an incident plane wave and calculating the field at the source dipole location.⁹⁶ The plane-wave calculation may be carried out by using a simple transverse equivalent network (TEN) model, which is a transmission-line model that represents the field behavior in the plane-wave problem. The PRS is assumed to be lossless and infinitesimally thin in the vertical direction, and therefore it is represented as a shunt susceptance B_L . Based on this simple model, formulas may be derived for the *beamwidth* and *pattern bandwidth* in the E - and H -planes, in terms of B_L .⁹⁶ The pattern bandwidth is defined from the frequency range over which the power density radiated at the angle θ_0 changes by less than 3 dB from the maximum value obtained at the center frequency (for which the structure is designed to radiate at angle θ_0).

Table 11-1 shows the beamwidth in the E - and H -planes for three separate cases: (1) broadside ($\theta_0 = 0^\circ$), (2) a general beam angle $0 < \theta_0 < 90^\circ$, and (3) end fire ($\theta_0 = 90^\circ$). The formulas are expressed in terms of the normalized shunt susceptance $\bar{B}_L = B_L \eta_0$, where η_0 is the intrinsic impedance of free space. It is seen that the beamwidths are equal at broadside, consistent with the fact that the beam is nearly omnidirectional. However, as the beam angle increases, the H -plane pattern becomes narrower, while the E -plane pattern becomes broader. A narrow beam can be obtained at the horizon in the H -plane (at least in theory) but not in the E -plane. It is noted that the beamwidth is inversely proportional to B_L^2 for a nonzero beam angle, but inversely proportional to B_L for a broadside beam. This means that the PRS must be much closer to a perfect reflecting surface in order to obtain a very narrow beam at broadside, compared with a nonzero beam angle.

TABLE 11-1 Expressions for Beamwidth

	E -plane	H -plane
Broadside	$2\sqrt{\frac{2n_1^3}{\pi \bar{B}_L^2}}$	$2\sqrt{\frac{2n_1^3}{\pi \bar{B}_L^2}}$
General beam angle	$\frac{2n_1^2 \sqrt{n_1^2 - \sin^2 \theta_0}}{\pi \bar{B}_L^2 \sin \theta_0 \cos^2 \theta_0}$	$\frac{2(\sqrt{n_1^2 - \sin^2 \theta_0})^3}{\pi \bar{B}_L^2 \sin \theta_0}$
End fire	Narrow beam not possible	$\frac{2(\sqrt{n_1^2 - 1})^3}{\pi \bar{B}_L^2}$

An analysis shows that the pattern bandwidth in the E - and H -planes is inversely proportional to B_L^2 for both a broadside beam and a conical beam. This means that for the same beamwidth the broadside pattern will have a much smaller pattern bandwidth than the conical beam.

REFERENCES

1. W. W. Hansen, Radiating Electromagnetic Waveguide (1940): U.S. Pat. 2,402,622.
2. L. O. Goldstone and A. A. Oliner, "Leaky-Wave Antennas—Part I: Rectangular Waveguides," *IRE Trans. Antennas Propagat.*, vol. AP-7 (October 1959): 307–319. See Section II for a simple discussion of the properties of leaky waves.
3. N. Marcuvitz, "On Field Representation in Terms of Leaky Modes or Eigenmodes," *IRE Trans. Antennas Propagat.*, vol. AP-4 (July 1956): 192–194.
4. T. Tamir and A. A. Oliner, "Guided Complex Waves—Part I: Fields at an Interface," *Proc. Inst. Elec. Eng. (London)*, vol. 110 (February 1963): 310–324.
5. T. Tamir and A. A. Oliner, "Guided Complex Waves—Part II: Relation to Radiation Patterns," *Proc. Inst. Elec. Eng. (London)*, vol. 110 (February 1963): 325–334.
6. C. J. Sletten and G. R. Forbes, "A New Antenna Radiator for VHF-UHF Communications," AFCRC-TR-57-114, Air Force Cambridge Research Center, Bedford, MA, 1957. See also the discussion in Chap. 10 of this handbook.
7. F. Schwing and S. T. Peng, "Design of Dielectric Grating Antennas for Millimeter Wave Applications," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-31 (February 1983): 199–209.
8. J. R. James and P. S. Hall, "Microstrip Antennas and Arrays—Part 2: New Array-Design Technique," *IEE J. Microwaves Optics Antennas*, no. 1 (1977): 175–181.
9. K. Solbach and B. Adelseck, "Dielectric Image Line Leaky-Wave Antennas for Broadside Radiation," *Electron. Lett.*, vol. 19 (August 1983): 640–644.
10. W. Rotman and S. J. Naumann, "The Design of Trough Waveguide Antenna Arrays," AFCRC-TR-58-154, Air Force Cambridge Research Center, Bedford, MA, 1958.
11. W. Rotman and A. A. Oliner, "Asymmetrical Trough Waveguide Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-7 (April 1959): 153–162.
12. C. H. Walter, *Traveling Wave Antennas* (New York: McGraw-Hill, 1965).
13. F. J. Zucker, "Surface- and Leaky-Wave Antennas," Chap. 16, *Antenna Engineering Handbook*, 1st Ed., H. Jasik (ed.) (New York: McGraw-Hill, 1961).
14. T. Tamir, "Leaky-Wave Antennas," Chap. 20, *Antenna Theory*, Part II, R. E. Collin and F. J. Zucker (eds.) (New York: McGraw-Hill, 1969).
15. A. A. Oliner and R. G. Malech, "Radiating Elements and Mutual Coupling," Chap. 2, *Microwave Scanning Antennas*, Vol. II, R. C. Hansen (ed.) (New York: Academic Press, Inc., 1966): 132–157.
16. C. H. Walter, *Traveling Wave Antennas*, Secs. 4 and 5 (New York: McGraw-Hill, 1965): 156–169.
17. L. O. Goldstone and A. A. Oliner, "Leaky-Wave Antennas—Part I: Rectangular Waveguides," *IRE Trans. Antennas Propagat.*, vol. AP-7 (October 1959): 307–319.
18. J. N. Hines, V. H. Rumsey, and C. H. Walter, "Traveling-Wave Slot Antennas," *Proc. IRE*, vol. 41 (November 1953): 1624–1631.
19. V. H. Rumsey, "Traveling Wave Slot Antennas," *J. Appl. Phys.*, vol. 24 (November 1953): 1358–1365.
20. A. L. Cullen, "On the Channel Section Waveguide Radiator," *Phil. Mag.*, vol. 40 (April 1949): 417–428.
21. W. Rotman, "The Channel Guide Antenna," *Proc. Natl. Electronics Conf.*, vol. 5 (1949): 190.
22. F. L. Whetten and C. A. Balanis, "Meandering Long Slot Leaky-Wave Waveguide Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-39 (November 1991): 1553–1560.

23. R. F. Harrington, "Propagation Along a Slotted Cylinder," *J. Appl. Phys.*, vol. 24 (November 1953): 1366–1371.
24. L. O. Goldstone and A. A. Oliner, "Leaky-Wave Antennas—Part II: Circular Waveguides," *IRE Trans. Antennas Propagat.*, vol. AP-9 (May 1961): 280–290.
25. J. N. Hines and J. R. Upson, "A Wide Aperture Tapered-Depth Scanning Antenna," Ohio State Univ. Res. Found., Report 667-7, Columbus, OH, December 1957.
26. R. C. Honey, "A Flush-Mounted Leaky Wave Antenna with Predictable Patterns," *IRE Trans. Antennas Propagat.*, vol. AP-7 (October 1959): 320–329.
27. D. D. King, "Dielectric Image Line," *J. Appl. Phys.*, vol. 23 (1952): 669.
28. H. W. Cooper, M. Hoffman, and S. Isaacson, "Image Line Surface Wave Antenna," *IRE Conv. Rec.*, Part 1 (1958): 230.
29. W. Rotman and N. Karas, "The Sandwich Wire Antenna: A New Type of Microwave Line Source Radiator," *IRE Conv. Rec.*, Part 1 (1957): 166.
30. W. Rotman and N. Karas, "The Sandwich Wire Antenna," *Microwave J.*, vol. 2, no. 8 (1959): 29.
31. T. Itoh and B. Adelseck, "Trapped Image Guide for Millimeter-Wave Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (December 1980): 1433–1436.
32. T. Rozzi and L. Ma, "Mode Completeness, Normalization, and Green's Function of the Inset Dielectric Guide," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-36 (March 1988): 542–551.
33. S. T. Peng and F. Schwing, "Dielectric Grating Antennas," R&D Tech. Rep., CORADCOM-78-3, Fort Monmouth, NJ, July 1978.
34. F. Schwing and S. T. Peng, "Design of Dielectric Grating Antennas for Millimeter-Wave Applications," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-31 (February 1983): 199–209.
35. F. Schwing and S. T. Peng, "Design of Periodically Corrugated Dielectric Antennas for Millimeter-Wave Applications," *Proc. 1982 Antenna Appl. Symp.* (1982).
36. K. L. Klohn, R. E. Horn, H. Jacobs, and E. Freibergs, "Silicon Waveguide Frequency Scanning Linear Array Antenna," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26 (October 1978): 764–773.
37. R. E. Horn, H. Jacobs, E. Freibergs, and K. L. Klohn, "Electronic Modulated Beam-Steering Silicon Waveguide Array Antenna," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (June 1980): 647–653.
38. R. E. Horn, H. Jacobs, K. L. Klohn, and E. Freibergs, "Single-Frequency Electronic-Modulated Analog Line Scanning Using a Dielectric Antenna," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-30 (May 1982): 816–820.
39. S. Kobayashi, R. Lampe, R. Mittra, and S. Ray, "Dielectric-Rod Leaky-Wave Antennas for Millimeter-Wave Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-29 (September 1981): 822–824.
40. T. Trinh, J. Malherbe, and R. Mittra, "A Metal-to Dielectric Waveguide Transition with Application to Millimeter-Wave Integrated Circuits," *Dig. Int. Microwave Symp.* (1980): 205–207.
41. T. N. Trinh, R. Mittra, and R. J. Paleta, "Horn Image-Guide Leaky-Wave Antenna," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29 (December 1981): 1310–1314.
42. T. Itoh and B. Adelseck, "Trapped Image-Guide Leaky-Wave Antennas for Millimeter-Wave Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-30 (May 1982): 505–509.
43. K. Solbach, "Slots in Dielectric Image Line as Mode Launchers and Circuit Elements," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29 (January 1981): 10–16.
44. S. T. Peng and T. Tamir, "TM Mode Perturbation Analysis of Dielectric Gratings," *Appl. Phys.*, vol. 7 (1975): 35.
45. T. Tamir and S. T. Peng, "Analysis and Design of Grating Couplers," *Appl. Phys.*, vol. 14 (1977): 235–254.
46. F. Schwing and A. A. Oliner, "Millimeter-Wave Antennas," Chap. 17, *Antenna Handbook: Theory, Applications and Design*, Y. T. Lo and S. W. Lee (eds.) (New York: Van Nostrand Reinhold, 1988): 17-48–17-82.
47. M. Guglielmi and A. A. Oliner, "A Practical Theory for Image Guide Leaky-Wave Antennas Loaded by Periodic Metal Strips," *Proc. 17th European Microwave Conf.* (1987): 549–554.

48. M. Guglielmi and A. A. Oliner, "Multimode Network Description of a Planar Periodic Metal-Strip Grating at a Dielectric Interface—Part I: Network Formulations," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-37 (March 1989): 534–541.
49. M. Guglielmi and A. A. Oliner, "Multimode Network Description of a Planar Periodic Metal-Strip Grating at a Dielectric Interface—Part II: Small-Aperture and Small-Obstacle Solutions," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-37 (March 1989): 542–552.
50. M. Guglielmi and H. Hochstadt, "Multimode Network Description of a Planar Periodic Metal-Strip Grating at a Dielectric Interface—Part III: Rigorous Solution," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-37 (May 1989): 902–909.
51. J. A. Encinar, "Mode-Matching and Point-Matching Techniques Applied to the Analysis of Metal-Strip-Loaded Dielectric Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-38 (September 1990): 1405–1412.
52. S. T. Peng, M. J. Shiau, A. A. Oliner, J. Borowick, W. Bayha, and F. Schwering, "A Simple Analysis Procedure for Dielectric Grating Antennas of Finite Width," 1984 IEEE AP-S Symp., Boston, MA, June 25–28, 1984.
53. J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna Theory and Design* (Stevenage, UK: Peter Peregrinus, Ltd., 1981).
54. J. R. James and P. S. Hall, *Handbook of Microstrip Antennas* (Stevenage, UK: Peter Peregrinus, Ltd., 1989).
55. A. G. Derneryd, "Linear Microstrip Array Antennas," Report TR 75057, Chalmers University of Technology, Division of Network Theory, 1975.
56. E. R. Cashen, R. Frost, and D. E. Yong, "Improvements Relating to Aerial Arrangements," British Provisional Patent (EMI, Ltd.), Specification 1294024, 1979.
57. M. Danielsen and R. Jorgensen, "Frequency Scanning Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-27 (March 1979): 146–151.
58. A. Henderson, A. E. England, and J. R. James, "New Low-Loss Millimeter-Wave Hybrid Microstrip Antenna Array," 11th European Microwave Conf., Amsterdam, The Netherlands, September 1981.
59. J. R. James and A. Henderson, "A Critical Review of Millimeter Planar Arrays for Military Applications," Military Microwaves Conf., London, England, October 20–22, 1982.
60. M. T. Brand and R. V. Gelsthorpe, "Experimental Millimetric Array Using Dielectric Radiators Fed by Means of Dielectric Waveguide," *Electron. Lett.*, vol. 17, no. 18 (September 1981): 633–635.
61. F. K. Schwering and A. A. Oliner, "Millimeter Wave Antennas Involving New Design Concepts," *Handbook of Optical and Microwave Components*, vol. I: *Microwave Passive and Antenna Components*, K. Chang (ed.) (New York: John Wiley and Sons, Inc., 1989): 710–731.
62. A. A. Oliner, "Recent Developments in Millimeter-Wave Antennas," *Alta Frequenza*, vol. LVIII, No. 5–6 (September–December 1989): 491–505.
63. A. A. Oliner, "Scannable Millimeter Wave Arrays," Final Report on RADC Contract No. F19628-84-K-0025, Polytechnic University, September 30, 1988, 528 pages in two volumes.
64. W. Rotman and N. Karas, "Some New Microwave Antenna Designs Based on the Trough Waveguide," *IRE Conv. Rec.*, Part 1 (1956): 230–235.
65. W. Rotman and A. A. Oliner, "Asymmetrical Trough Waveguide Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-7 (April 1959): 153–162.
66. T. Yoneyama and S. Nishida, "Nonradiative Dielectric Waveguide for Millimeter-Wave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29 (November 1981): 1188–1192.
67. A. Sanchez and A. A. Oliner, "A New Leaky Waveguide for Millimeter Waves Using Nonradiative Dielectric (NRD) Waveguide—Part I: Accurate Theory," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35 (August 1987): 737–747.
68. T. Yoneyama, Letter to A. A. Oliner, July 4, 1983.
69. Q. Han, A. A. Oliner, and A. Sanchez, "A New Leaky Waveguide for Millimeter Waves Using Nonradiative Dielectric (NRD) Waveguide—Part II: Comparison with Experiments," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35 (August 1987): 748–752.

70. A. A. Oliner, S. T. Peng, and K. M. Sheng, "Leakage from a Gap in NRD Guide," *Dig. 1985 IEEE Intl. Microwave Symp.* (1985): 619–622.
71. H. Shigesawa, M. Tsuji, and A. A. Oliner, "Coupling Effects in an NRD Guide Leaky Wave Antenna," *Dig. Natl. Radio Sci. Meeting* (1986): 27.
72. T. Yoneyama, T. Kuwahara, and S. Nishida, "Experimental Study of Nonradiative Dielectric Waveguide Leaky Wave Antenna," *Proc. 1985 Int. Symp. Antennas Propagat.* (August 1985).
73. A. A. Oliner and P. Lampariello, "A New Simple Leaky Wave Antenna for Millimeter Waves," *Dig. 1985 North American Radio Sci. Meeting* (1985): 57.
74. P. Lampariello, F. Frezza, H. Shigesawa, M. Tsuji, and A. A. Oliner, "Guidance and Leakage Properties of Offset Groove Guide," *Dig. IEEE Int. Microwave Symp.* (1987): 731–734.
75. H. Shigesawa, M. Tsuji, and A. A. Oliner, "Theoretical and Experimental Study of an Offset Groove Guide Leaky Wave Antenna," *Dig. IEEE AP-S Int. Symp.* (1987): 628–632.
76. P. Lampariello, F. Frezza, H. Shigesawa, M. Tsuji, and A. A. Oliner, "A Versatile Leaky-Wave Antenna Based on Stub-Loaded Rectangular Waveguide: Part I—Theory," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-46 (July 1998): 1032–1041.
77. F. Frezza, P. Lampariello, H. Shigesawa, M. Tsuji, and A. A. Oliner, "A Versatile Leaky-Wave Antenna Based on Stub-Loaded Rectangular Waveguide: Part II—Effects of Flanges and Finite Stub Length," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-46 (July 1998): 1042–1046.
78. M. Tsuji, H. Shigesawa, F. Frezza, P. Lampariello, and A. A. Oliner, "A Versatile Leaky-Wave Antenna Based on Stub-Loaded Rectangular Waveguide: Part III—Comparisons with Measurements," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-46 (July 1998): 1047–1055.
79. P. Lampariello and A. A. Oliner, "New Equivalent Networks with Simple Closed-Form Expressions for Open and Slit-Coupled E-Plane Tee Junctions," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-41 (May 1993): 839–847.
80. F. Frezza, M. Guglielmi, and P. Lampariello, "Millimetre-Wave Leaky-Wave Antennas Based on Slitted Asymmetric Ridge Waveguides," *IEE Proc., Microw. Antennas Propagat.*, vol. 141 (June 1994): 175–180.
81. A. A. Oliner and S. J. Xu, "A Novel Phased Array of Leaky-Wave NRD Guides," *Dig. Natl. Radio Sci. Meeting* (1987): 139.
82. P. Lampariello and A. A. Oliner, "A Novel Phased Array of Printed-Circuit Leaky-Wave Line Sources," *Proc. 17th European Microwave Conf.* (1987): 555–560.
83. M. Guglielmi and A. A. Oliner, "A Linear Phased Array of Periodic Printed-Circuit Leaky-Wave Line Sources that Permits Wide Two-Dimensional Scan Coverage," *Dig. Natl. Radio Sci. Meeting* (1988): 420.
84. A. Ip and D. R. Jackson, "Radiation from Cylindrical Leaky Waves," *IEEE Trans. Antennas Propagat.*, vol. 38 (April 1990): 482–488.
85. G. von Trentini, "Partially Reflecting Sheet Arrays," *IEEE Trans. Antennas Propagat.*, vol. 4 (October 1956): 666–671.
86. A. P. Feresidis and J. C. Vardaxoglou, "High Gain Planar Antenna Using Optimised Partially Reflective Surfaces," *IEE Proc. Microwave Antennas Propagat.*, vol. 148 (December 2001): 345–350.
87. N. G. Alexopoulos and D. R. Jackson, "Fundamental Superstrate (Cover) Effects on Printed Circuit Antennas," *IEEE Trans. Antennas Propagat.*, vol. 32 (August 1984): 807–816.
88. D. R. Jackson and N. G. Alexopoulos, "Gain Enhancement Methods for Printed Circuit Antennas," *IEEE Trans. Antennas Propagat.*, vol. 33 (September 1985): 976–987.
89. D. R. Jackson and A. A. Oliner, "A Leaky-Wave Analysis of the High-Gain Printed Antenna Configuration," *IEEE Trans. Antennas Propagat.*, vol. 36 (July 1988): 905–910.
90. H. Y. Yang and N. G. Alexopoulos, "Gain Enhancement Methods for Printed Circuit Antennas," *IEEE Trans. Antennas Propagat.*, vol. 35 (July 1987): 860–863.
91. D. R. Jackson, A. A. Oliner, and A. Ip, "Leaky-Wave Propagation and Radiation for a Narrow-Beam Multiple-Layer Dielectric Structure," *IEEE Trans. Antennas Propagat.*, vol. 41 (March 1993): 344–348.

92. M. Thevenot, C. Cheype, A. Reineix, and B. Jecko, "Directive Photonic-Bandgap Antennas," *IEEE Trans. Microwave Theory Tech.*, vol. 47 (November 1999): 2115–2122.
93. T. Zhao, D. R. Jackson, J. T. Williams, H. Y. Yang, and A. A. Oliner, "2-D Periodic Leaky-Wave Antennas—Part I: Metal Patch Design," *IEEE Trans. Antennas Propagat.*, vol. 53 (November 2005): 3505–3514.
94. T. Zhao, D. R. Jackson, and J. T. Williams, "2-D Periodic Leaky-Wave Antennas—Part II: Slot Design," *IEEE Trans. Antennas Propagat.*, vol. 53 (November 2005): 3515–3524.
95. G. Lovat, P. Burghignoli, F. Capolino, D. R. Jackson, and D. R. Wilton, "High-Gain Omnidirectional Radiation Patterns from a Metal Strip Grating Leaky-Wave Antenna," *IEEE Intl. Antennas and Propagation Symp.*, Honolulu, HI, July 2007.
96. T. Zhao, D. R. Jackson, and J. T. Williams, "General Formulas for 2D Leaky Wave Antennas," *IEEE Trans. Antennas Propagat.*, vol. 53 (November 2005): 3525–3533.

Chapter 12

Helical Antennas

Howard E. King
Jimmy L. Wong

The Aerospace Corporation
with additions by

Edward H. Newman

The Ohio State University

CONTENTS

12.1	INTRODUCTION.....	12-2
12.2	AXIAL-MODE HELICAL ANTENNAS.....	12-3
12.3	ARRAY OF HELICAL ANTENNAS	12-16
12.4	NORMAL-MODE HELICAL ANTENNAS	12-18
12.5	OTHER HELICAL ANTENNAS	12-19
12.6	SATELLITE WIDEBAND HELICAL ANTENNA DESIGN	12-22
12.7	RECENT DEVELOPMENTS	12-22

12.1 INTRODUCTION

Helical antennas consist of a single conductor or multiple conductors wound into a helical shape. Although a helix can radiate in many modes, the axial mode and the normal mode are the ones of general interest. The axial mode, the most commonly used mode, provides maximum radiation along the helix axis, which occurs when the helix circumference is of the order of one wavelength. The normal mode, which yields radiation broadside to the helix axis, occurs when the helix diameter is small with respect to a wavelength. Higher-order-radiation modes are also possible; for example when the helix dimensions exceed those required for the axial mode, a conical or multilobed pattern will result, as illustrated in Figure 12-1.

The basic concepts of a helix antenna were established by Kraus^{1,2} in 1947, and much of Kraus's early results were summarized by Harris.³ Generally, helical antennas are wound with a single conduction. However, a helix can be designed with bifilar,⁴ quadrifilar,^{5,6} or multifilar⁷ windings. Radiation characteristics of bifilar helices operating in the backfire mode have been described by Patton.⁸ An advantage of a backfire helix is that it does not generally require a ground plane.

The helix-antenna parameters are defined as follows (see Figure 12-2):

D = diameter of helix (center to center)

C = circumference of helix = πD

S = spacing between turns (centers to center)

α = pitch angle = $\tan^{-1}(S/\pi D)$

N = number of turns

L = axial length of helix = NS

d = diameter of helix conductor

ℓ = length of one turn = $\sqrt{(\pi D)^2 + S^2}$

Although helical antennas are normally constructed with a circular cross section, elliptical helical antennas have also been investigated.⁹

This chapter presents detailed helical-antenna design information. Measured impedance, pattern, gain, and axial-ratio characteristics are shown for a variety of helical-antenna configurations. Empirical relations that express the antenna radiation performance characteristics as a function of wavelength and the helix design parameters (diameter, pitch angle, and number of turns) are derived on the basis of measured data.

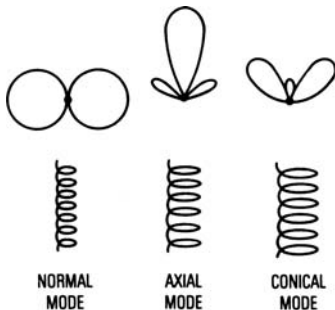


FIGURE 12-1 Three radiation modes of a helical antenna

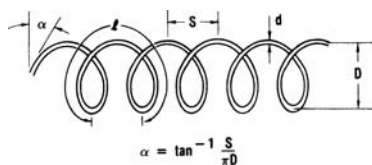


FIGURE 12-2 Helix geometry

12.2 AXIAL-MODE HELICAL ANTENNAS

The helical beam antenna is a very simple structure possessing a number of interesting properties including wideband impedance characteristics and circularly polarized radiation. It requires a simple feed network, and its radiation characteristics are reasonably predictable. Either right-hand or left-hand circular polarization may be generated by a helical beam antenna. A helix wound like a right-hand screw radiates or receives right-hand circular polarization, while a helix wound like a left-hand screw radiates or receives left-hand circular polarization.

Helical antennas are generally constructed with a uniform diameter and operated in conjunction with a ground plane, cavity, or helical launcher.¹⁰ However, as will be shown, nonuniform-diameter helical structures can be employed to widen the bandwidth of a uniform-diameter helical antenna and to improve radiation performance characteristics.^{11,12}

A typical uniform-helix configuration is shown in Figure 12-3. The helix is backed by a circular cavity, rather than a conventional ground plane, to reduce the back radiation and enhance the forward gain. Although the helix of Figure 12-3 was designed¹¹ specifically to operate in the ultrahigh-frequency (UHF) range from about 650 to 1100 MHz, the helix and the cavity dimensions can be scaled to other frequencies of operation as well. The helix is fed by a coaxial connection at the bottom of the cavity. The helical conductor can be made from round tubing or flat strip and supported with a lightweight foam dielectric cylinder or by radial dielectric rods. The center mechanical support can be constructed from either metal or dielectric material.

Impedance and VSWR

The impedance characteristics of a helical antenna have been established by Glasser and Kraus.¹³ When the helix circumference is less than two-thirds wavelength, the terminal impedance is highly sensitive to frequency changes. However, when the circumference is of the order of one wavelength (axial mode), the terminal impedance is nearly a pure resistance and is given approximately (within ± 20 percent) by the empirical relation $R = 140 C/\lambda \Omega$. The relatively constant terminal impedance may be explained by the rapidly attenuating character of the total outgoing wave near the input end and the total reflected wave near the open end.¹⁴

For the UHF helix configuration of Figure 12-3, a 4.7-in- (119-mm-) long, linear-taper impedance transformer is used to match the 140- Ω helix to the 50- Ω coaxial line. The transformer is a microstripline constructed from Teflon-fiberglass printed-circuit board. The same transformer can be used for all axial-mode helices—uniform-, tapered-, or nonuniform-diameter. Another approach to attain a 50- Ω impedance for helical beam antennas has been described by Kraus.¹⁵

Figure 12-4 depicts the impedance characteristics of a five-turn helix with and without the impedance-matching transformer, and Figure 12-5 depicts the voltage-standing-wave-ratio (VSWR) response measured at the input of the matching transformer for a seven-turn uniform helix and also for the same helix with two additional turns of tapered diameter, as illustrated by the inset. By adding the two-turn end taper, a significant reduction in VSWR over a wide frequency

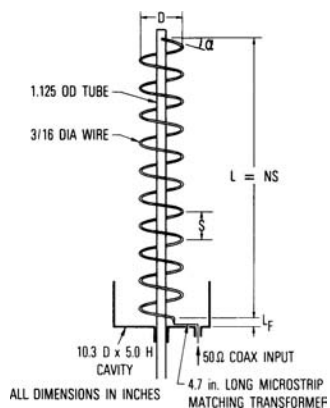


FIGURE 12-3 Mechanical arrangement of a cavity-backed helix for 650 to 1100 MHz

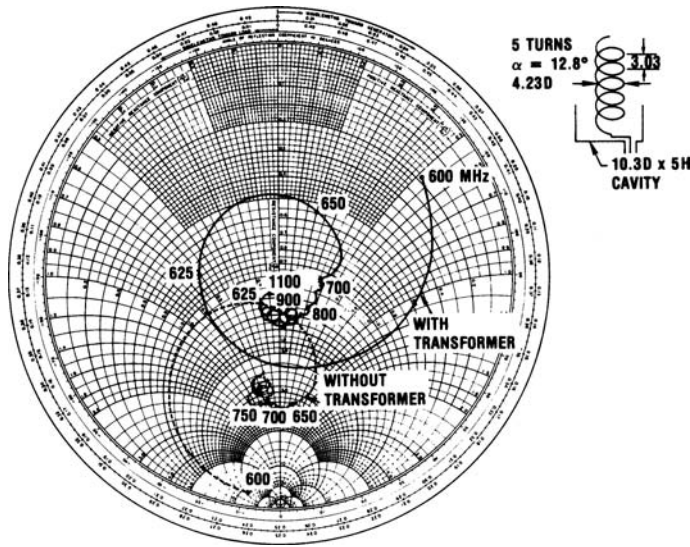


FIGURE 12-4 Impedance characteristics of a five-turn helix with and without an impedance-matching transformer

band can be achieved (dashed curve). The reduction in VSWR is due to the suppression of the reflected currents by the tapered end section.^{16,17} Also, it is noted that the low-frequency characteristics are essentially unchanged with the cutoff at ~ 534 MHz, corresponding to $C/\lambda \sim 0.75$, where C is the circumference of the 5.28-in- (134.1-mm-) diameter helix. The low-frequency cutoff characteristics agree well with theoretical predictions.^{1,2} The VSWR characteristics for longer helices are generally similar to those of Figure 12-5.

Uniform-Diameter Helices

As a first approximation, the radiation pattern of an axial-mode helix may be obtained by assuming a single traveling wave of uniform amplitude along the conductor. By the principle

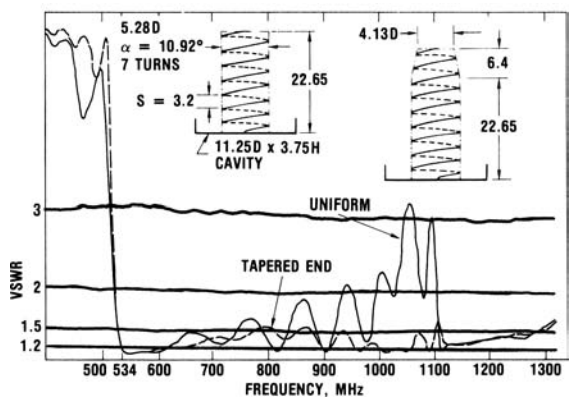


FIGURE 12-5 VSWR of a seven-turn uniform helix and the same helix with two additional turns wound on a tapered diameter

of pattern multiplication, the far-field pattern is the product of the pattern of one turn and the pattern of an array of N isotropic elements with spacing S , where N equals the number of turns and S is the spacing between turns. If the pattern of a single turn is approximated by $\cos \theta$, where θ is the angle measured from the axis of the helix, the total radiation pattern becomes

$$E(\theta) = A \frac{\sin(N\psi/2)}{\sin(\psi/2)} \cos \theta \quad (12-1)$$

where

A = normalization factor

$$\psi = \frac{2\pi}{\lambda} S \cos \theta - \delta \quad (12-2)$$

$\delta = \frac{2\pi}{\lambda} \frac{\ell}{v/c}$ = progressive phase between turns

ℓ = length of one turn

v = phase velocity along helical conductor

c = velocity of light in free space

Kraus² has shown that for an axial-mode helix the relative phase velocity, $p = v/c$, of the wave propagating along the helical conductor is in close agreement with that required to satisfy the Hansen-Woodyard condition for an end-fire array with increased directivity¹⁸; i.e.,

$$p = \frac{(\ell / \lambda)}{(S / \lambda) + (2N + 1) / 2N} \quad (12-3)$$

Thus, for the increased-directivity condition, the quantity ψ may be written as

$$\psi = \frac{2\pi}{\lambda} S (\cos \theta - 1) - \frac{\pi}{N} \quad (12-4)$$

and

$$A = \sin(\pi / 2N) \quad (12-5)$$

The radiation is elliptically polarized. The ellipticity ratio or axial ratio (AR) of an N -turn helix operating in the axial mode for the increased-directivity condition is given by

$$\text{AR} = \frac{2N + 1}{2N} \quad (12-6)$$

If N is large, the axial ratio approaches unity and the polarization is nearly circular. In a practical situation, the observed axial ratio may be expected to deviate somewhat from the theoretical value for various reasons such as methods of construction, tolerances, instrumentation, and range errors. For a uniform helix consisting of at least a few turns, an axial ratio of the order of 1 dB is not uncommon.

To provide parametric design equations for axial-mode helices, Kraus² has suggested the following relations for the half-power beamwidth (HPBW) and gain (G) as a function of C/λ and NS/λ for constant-pitch helices with $12^\circ < \alpha < 15^\circ$, $3/4 < C/\lambda < 4/3$, and $N > 3$:

$$\text{HPBW} = K_B / (C / \lambda) \sqrt{NS / \lambda} \quad (12-7)$$

$$G = K_G (C/\lambda)^2 (NS/\lambda) \quad (12-8)$$

where K_B is the HPBW factor, K_G is the gain factor, $C = \pi D$ is the circumference, and NS is the axial length. On the basis of a number of pattern measurements on helices < 10 turns, Kraus quasi-empirically established $K_B = 52$. Also, he derived $K_G = 15$ for the directive gain (lossless antenna) based on the approximation $G = 41,250/(\text{HPBW})^2$, where HPBW is in degrees.

A gain-beamwidth product $G(\text{HPBW})^2 < 41,250$ is generally expected for most practical antennas¹⁹ because of minor-lobe radiation and beam-shape variations.

Additional data on the radiation characteristics of axial-mode helices are available in the literature.^{11,20-22} The low-frequency and high-frequency limits were investigated by Maclean and Kouyoumjian²⁰ and by Maclean.²³ In a practical situation, the pattern and gain versus frequency characteristics are of interest to the antenna designer. Such data would allow the designer to optimize the helix parameter for operation over a specified bandwidth.

The discussions that follow summarize the results of an extensive study of the gain and pattern characteristics of uniform helical antennas, one or eight wavelengths long, in the UHF range from about 650 to 1100 MHz. With reference to Figure 12-3, the helices were constructed by winding $\frac{3}{16}$ -in-diameter copper tubing about a Styrofoam cylindrical form that was concentric with 1.125-in-diameter metallic support tube. The helix dimensions can be scaled to other frequencies of operation with essentially the same radio-frequency performance characteristics. All gain and pattern measurements were made with respect to the phase center of the helix, which was estimated to be one-fourth of the helix length from the feed point.²⁴

Fixed-Length Helices Figures 12-6 and 12-7 show the gain and HPBW versus frequency characteristics, respectively, of a 30-in- (762-mm-) long ($NS = 30$ in) and 4.3-in- (109.22-mm-) diameter helix for three helix pitch angles ($\alpha = 12.5^\circ, 13.5^\circ$, and 14.5°). The helix with a smaller pitch angle (more turns per unit length) yields a higher peak gain and a lower cutoff frequency. With $N = 8.6$ to 10 turns, it appears that the gain-frequency slope is approximately proportional to f^3 and the HPBW-frequency slope is approximately proportional to $f^{-3/2}$, where f equals the frequency, which are in general agreement with Kraus for $C/\lambda < 1.1$ (see Eqs. 12-7 and 12-8). However, as will be shown later, experimental data indicate that the gain-frequency slope depends on the antenna length and is approximately proportional to $f^{\sqrt{N}}$.

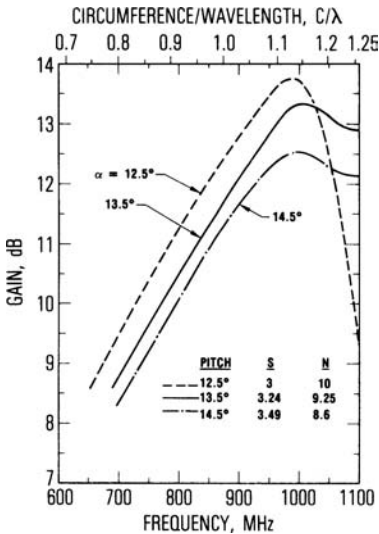


FIGURE 12-6 Gain versus frequency of a fixed-length ($NS = 30$ in, or 762 mm) helix; $\alpha = 12.5^\circ, 13.5^\circ$, and 14.5° ; $D = 4.3$ in (109 mm)

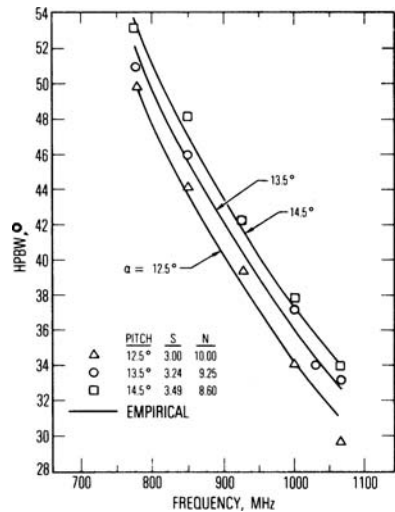


FIGURE 12-7 Half-power beamwidth versus frequency of a fixed-length helix ($NS = 30$ in, or 762 mm); $\alpha = 12.5^\circ, 13.5^\circ$, and 14.5° ; $D = 4.3$ in (109 mm)

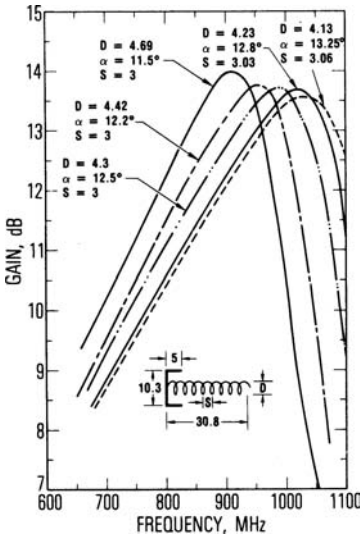


FIGURE 12-8 Gain of a fixed-length ($NS = 30$ in, or 762 mm) helical antenna with different diameters

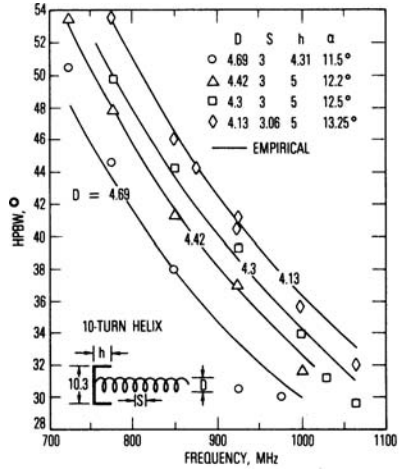


FIGURE 12-9 Half-power beamwidths of a 10-turn fixed-length ($NS = 30$ in, or 762 mm) helix with different diameters

Figures 12-8 and 12-9 show the gain and HPBW characteristics, respectively, of a 30-in-long ($N \approx 10$ turns) helix with variable diameter and pitch angle. The peak gain varies by less than 0.5 dB. In general, a slightly higher peak gain is observed for a larger-diameter helix with a smaller pitch angle, but the bandwidth is narrower than that of a smaller-diameter helix with a larger pitch angle.

The axial ratio, a measure of the purity of the circularly polarized wave, is generally less than 1.5 dB for $0.8 < C/\lambda < 1.2$. The axial ratio can be improved by tapering the last two turns of the helix, particularly at the high end of the band.^{12,17,25}

Variable-Length Helices Figures 12-10 and 12-11 show the gain and HPBW versus frequency, respectively, of helices consisting of 5 to 35 turns with constant pitch ($\alpha = 12.8^\circ$, $S = 3.03$ in, or 76.96 mm) and constant diameter ($D = 4.23$ in, or 107.4 mm). These helices were designed to operate over the UHF test frequencies with the helix circumference varied from 0.75λ to 1.25λ . N was selected as 5, 10, 12, 15, 18, 22, 26, 30, and 35 turns. The gain is referred to as a circularly polarized illuminating source. The gain curves reveal that the peak gain occurs at $C/\lambda = 1.55$ for $N = 5$ and at a lower value, $C/\lambda = 1.07$, for $N = 35$. The gain-frequency slope is not proportional to f^3 for all values of N (see Eq. 12-8); e.g., for $N = 5$ the gain varies approximately as $f^{2.5}$, and for $N = 35$ the gain follows approximately an f^6 slope, where f is the frequency.

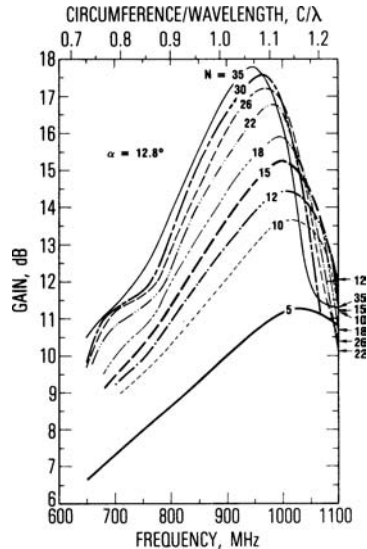


FIGURE 12-10 Gain versus frequency for a 5- to 35-turn helix; $\alpha = 12.8^\circ$, $D = 4.23$ in (107.4 mm)

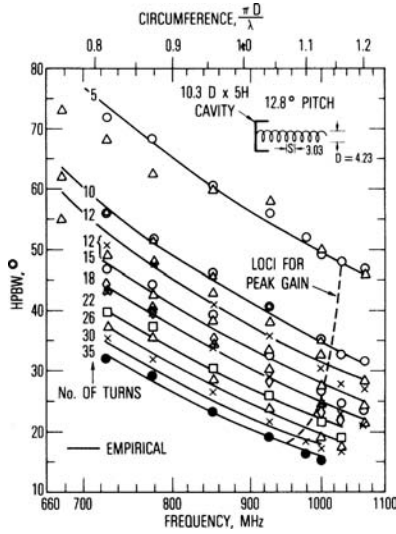


FIGURE 12-11 Half-power beamwidths versus frequency for a 5- to 35-turn helix; $\alpha = 12.8^\circ$, $D = 4.23$ in (107.4 mm)

Typical measured radiation patterns for $N = 5, 10, 18$, and 35 are shown in Figure 12-12. The axial ratio is ~ 1 dB over most of the measurement frequency range and is slightly higher at the band edges. The HPBWs are generally within $\pm 1^\circ$ in the two orthogonal principal planes. At frequencies a few percent above the peak gain frequency, the patterns begin to deteriorate. The beamwidth broadens rapidly, and the first sidelobes merge with the main lobe as the operating frequency approaches the upper limit.

Parametric helix characteristic curves—gain and halfpower beamwidth—for the 4.23-in- (107.4-mm-) diameter constant-pitch ($\alpha = 12.8^\circ$) helix are shown in Figures 12-13 and 12-14, respectively, as a function of axial length NS/λ with circumference $\pi D/\lambda$ as a parameter. Thus, for a specified length and diameter, the helix gain and HPBW can be estimated.

Figure 12-15 depicts the gain-beamwidth product $K = G (\text{HPBW})^2$ based on the measured gain data of Figure 12-10 and the HPBW data of Figure 12-11. The quantity K is useful for estimating the gain when the HPBW is known, and vice versa. It should be noted, however, that the gain-beamwidth product is not constant but depends on N and frequency. All curves have been smoothed to within ± 5 percent of the measured-data points.

Empirical Relations On the basis of the gain data of Figures 12-6, 12-8, and 12-10, the peak gain may be empirically expressed as¹¹

$$G_p = 8.3 \left(\frac{\pi D}{\lambda_p} \right)^{\sqrt{N+2}-1} \left(\frac{NS}{\lambda_p} \right)^{0.8} \left[\frac{\tan 12.5^\circ}{\tan \alpha} \right]^{\sqrt{N}/2} \quad (12-9)$$

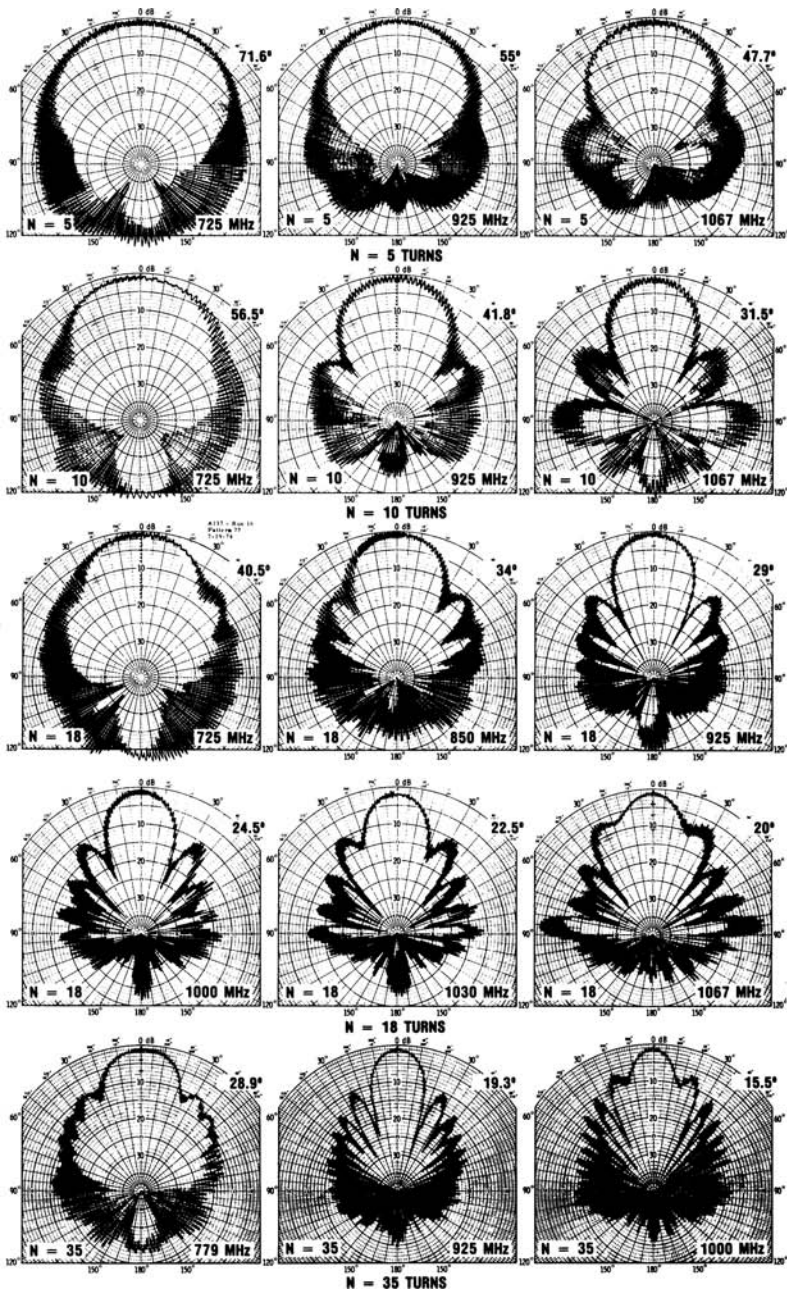


FIGURE 12-12 Typical radiation patterns of a 5- to 35-turn helix; $\alpha = 12.8^\circ$, $D = 4.23$ in (107.4 mm)

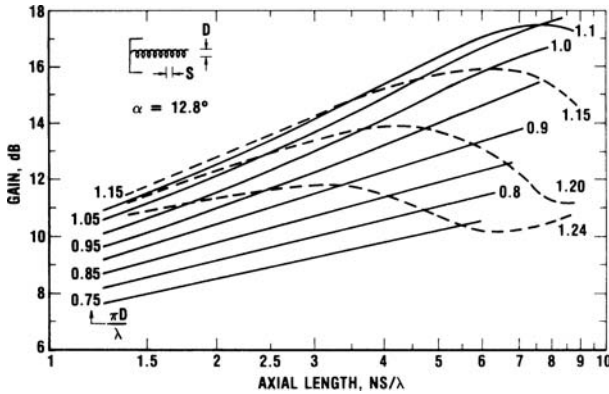


FIGURE 12-13 Parametric helix-antenna gain curves as a function of axial length with circumference as a parameter

where λ_p is the wavelength at peak gain. The computed values for the fixed-length helices ($NS = \text{constant}$) are within ± 0.1 dB of the measured data, as depicted in Figure 12-16. The data points indicated by the circle were obtained by varying the diameter and pitch angle while keeping the length constant with $N = 10$ turns, and those indicated by the triangle were obtained by varying the pitch angle while keeping the length and diameter constant ($N \approx 8.6$ to 10). The diameters of the various experimental helices are shown on the top of the figure. For the fixed-length helices, the peak gains occur at nearly the same circumference, $\pi D/\lambda \sim 1.135$.

Figure 12-17 is a plot of the peak gain versus N for a constant-pitch ($\alpha = 12.8^\circ$) helix. The corresponding values of C/λ_p are also shown in the figure, where λ_p is the wavelength at peak gain. The peak gain is not quite proportional to the number of turns; i.e., doubling the number of turns does not yield a 3-dB increase in the peak gain. The computed values for the peak gain, using Eq. 12-9 with $\alpha = 12.8^\circ$ and $N = 5$ to 35 turns, are within ± 0.1 dB of the measured data.

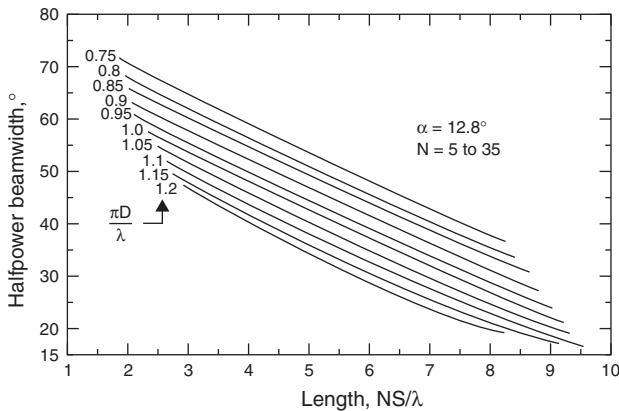


FIGURE 12-14 Parametric helix-antenna half-power-beamwidth curves as a function of axial length with circumference as a parameter

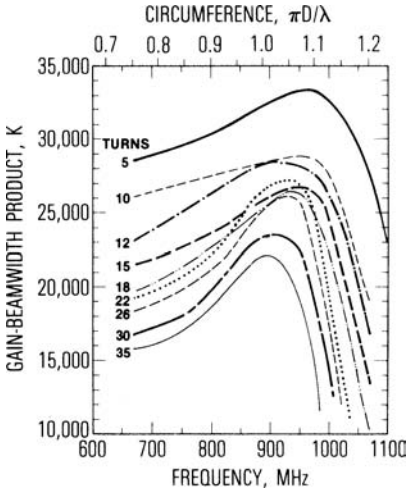


FIGURE 12-15 Gain-beamwidth products of a 5- to 35-turn helix; $\alpha = 12.8^\circ$, $D = 4.23$ in (107.4 mm)

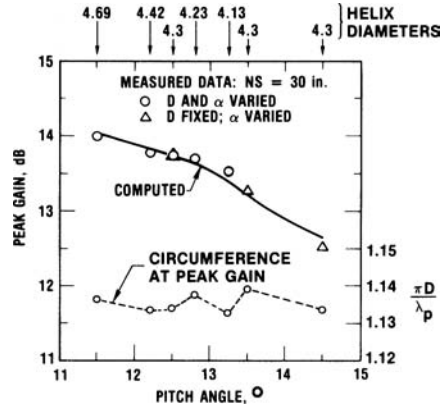


FIGURE 12-16 Peak gain of a fixed-length helix ($NS = 30$ in, or 762 mm) as a function of a pitch angle

A similar empirical expression for the HPBW is as follows:²⁶

$$\text{HPBW} \approx \frac{K_B \left(\frac{2N}{N+5} \right)^{0.6}}{\left(\frac{\pi D}{\lambda} \right)^{\sqrt{N}/4} \left(\frac{NS}{A} \right)^{0.7}} \left(\frac{\tan \alpha}{\tan 12.5^\circ} \right)^{\sqrt{N}/4} \quad (12-10)$$

where K_B is a constant in degrees. For the helices constructed by using the arrangement of Figure 12-3, it was found that, with $K_B \approx 61.5^\circ$, Eq. 12-10 matches the measured data within \pm a few percent over the useful operating frequency range of the helix, as shown in Figures 12-7, 12-9, and 12-11. For helices that employ a different construction technique (e.g., tape helices that are wound on a dielectric support instead of using a metallic central support rod), Eq. 12-10 can still be applied, but a slightly different value of K_B must be used. The measured HPBWs for the helices investigated are generally 10 to 20 percent

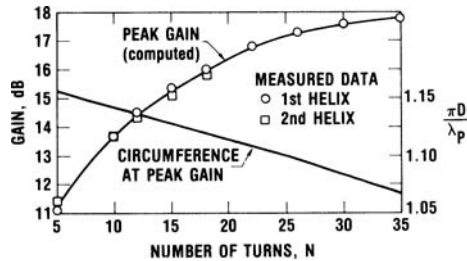


FIGURE 12-17 Peak gain of a variable-length helix with $\alpha = 12.8^\circ$

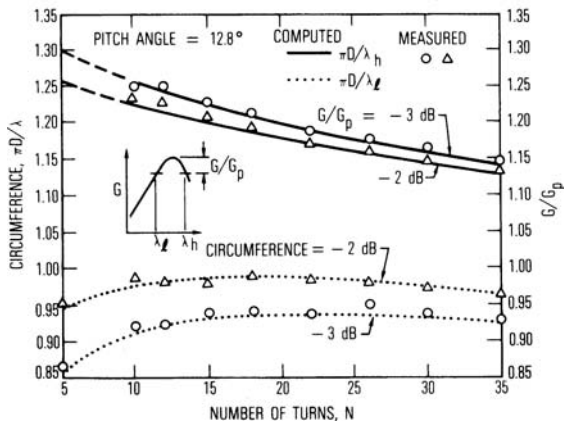


FIGURE 12-18 Gain-bandwidth characteristics of a uniform helix with $\alpha = 12.8^\circ$

wider than Kraus's formula (see Eq. 12-7). It should be mentioned that Eqs. 12-9 and 12-10 are not unique; however, these relations are useful as a design tool.

The helix bandwidth may be defined as the operating frequency range over which the gain drops by an allowable amount. The -3 -dB and -2 -dB bandwidths as a function of N may be obtained by using the curves of Figure 12-18. For example, if the -3 -dB or -2 -dB bandwidth is desired for a given N , you determine the values of $\pi D/\lambda_h$ and $\pi D/\lambda_\ell$ from Figure 12-18, where λ_h and λ_ℓ are the free-space wavelengths corresponding to the upper frequency limit f_h and the lower frequency limit f_ℓ , respectively. The choice of a -3 -dB or -2 -dB bandwidth depends upon the antenna designer's application. The frequency limits f_h and f_ℓ may be determined by using the measured gain data of Figure 12-10. Note that the gain varies approximately as $f^{\sqrt{N}}$ for $f < f_p/1.04$ and as $f^{-3\sqrt{N}}$ for $f > 1.03 f_p$, where f_p is the frequency at peak gain. On the basis of these observations, the bandwidth frequency ratio may be empirically expressed as¹¹

$$\frac{f_h}{f_\ell} \approx 1.07 \left(\frac{0.91}{G/G_p} \right)^{4/(3\sqrt{N})} \quad (12-11)$$

where G_p is the peak gain from Eq. 12-9. The computed bandwidth characteristics for $G/G_p = -3$ dB and -2 dB agree reasonably well with the measured data as shown in Figure 12-18. The bandwidth decreases as the axial length of the helix increases. This bandwidth behavior follows the same trends described by Maclean and Kouyoumjian,²⁰ although these authors employ a sidelobe criterion rather than a gain criterion. Beyond the $G/G_p = -3$ dB point, the gain drops off sharply at the high-frequency end as the upper limit for the axial mode is approached.

Nonuniform-Diameter and Tapered Helices

The tapered-diameter^{27,28} and nonuniform-diameter¹² helices represent additional types of axial mode helical antennas. These helical configurations are capable of providing a wider bandwidth than a conventional uniform (constant-diameter) helix. A nonuniform helix

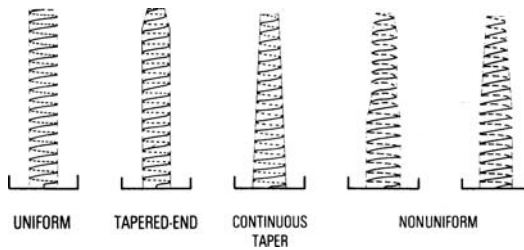


FIGURE 12-19 Various axial-mode helical configurations

consists of multiple uniform-diameter helical sections that are joined together by short tapered transitions. Some studies have been made to investigate the broadband-frequency response of nonuniform-diameter helical antennas.^{29,30} The various types of helical-antenna configurations (uniform, tapered, and nonuniform-diameter) are shown in Figure 12-19, and the radiation performance characteristics are described in Wong and King.¹²

The experimental helices were wound with thin copper strips 0.468 in (11.887 mm) wide. The plane of the strip (the wide dimensions of the strip) was wound orthogonally to the helix axis similarly to a Slinky toy. Helices wound with round conductors or with metallic tapes (wound so that the plane of the tape is parallel to the helix axis) yield similar results. A constant-pitch spacing of 3.2 in (81.28 mm) was selected, and the helix was backed by an 11.25-in- (285.75-mm-) diameter \times 3.75-in- (95.25-mm-) high cavity. The feed arrangement and the metallic central support tube are similar to those of Figure 12-3. The VSWR of a nonuniform helix is $\leq 1.5:1$ over the test frequency range of 650 to 1100 MHz (similar to the dashed curve of Figure 12-5).

Tapered-End Helices For reference purposes, the gain and axial-ratio characteristics of an 18-turn uniformly wound helix, with $D = 4.59$ in (116.59 mm) and $\alpha = 12.5^\circ$, are shown in Figure 12-20. Also shown in this figure are the gain and axial-ratio characteristics of a similar helix with the same overall length, but the diameter of the last two turns is tapered from 4.59 in to 2.98 in (75.69 mm). It can be seen that the end taper provides a marked improvement in axial-ratio characteristics,^{12,16,17} although the peak gain is reduced slightly. Figure 12-21 compares the radiation patterns of a uniform and a tapered-end helix measured at the same frequency. The HPBW's are approximately the same, but the axial-ratio improvement is observed over the entire pattern (on axis and off axis). Similar results have been reported by Donn.²⁵

Continuously Tapered Helices The gain and axial characteristics of a continuously tapered helix,¹² often referred to as a *conical helix*, are shown in Figure 12-22. The helix consists of 17.64 turns with a constant-pitch spacing of

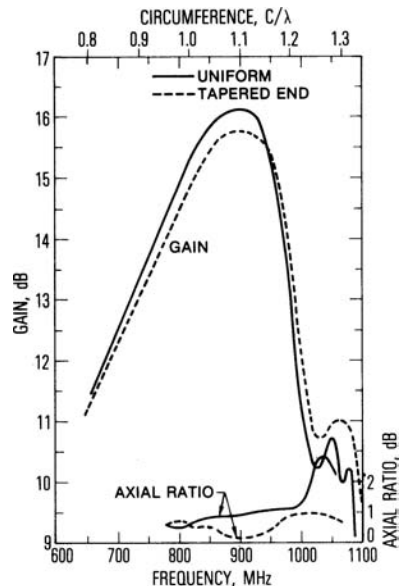


FIGURE 12-20 Gain and axial-ratio characteristics of an 18-turn uniformly wound and tapered-end helix

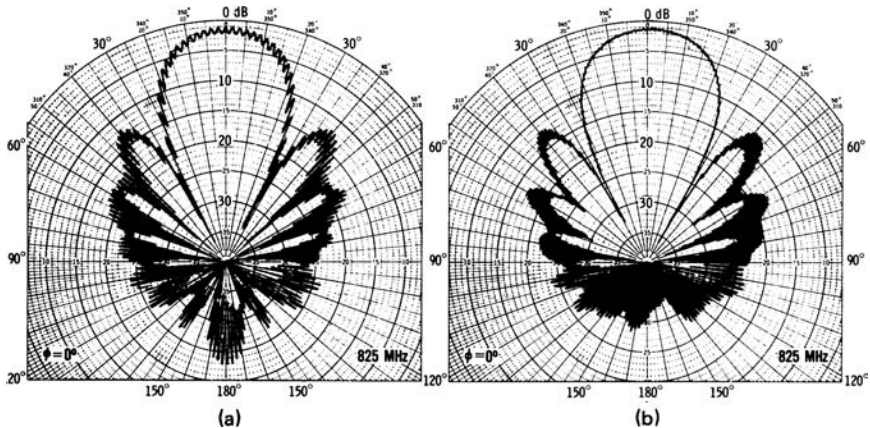


FIGURE 12-21 Radiation patterns of an 18-turn (a) uniform and (b) tapered-end helix

3.2 in (81.28 mm) and tapers from a 5.32-in (135.12-mm) diameter at the base to a 2.98-in (75.69-mm) diameter at the top (see inset). Typical measured patterns are shown in Figure 12-23. The peak gain is slightly lower than that of the uniform helix, but the pattern characteristics and bandwidth performance are much better.

Nonuniform-Diameter Helix A nonuniform helix provides a unique approach for widening the bandwidth of a helical antenna with improved gain and pattern characteristics. It may be constructed with two or more uniform helix sections of different diameters or a combination of uniform and tapered sections. The dimensions of the different helical sections (diameter, number of turns, etc.) can be varied to synthesize an antenna with a specific gain-frequency response.^{12,31}

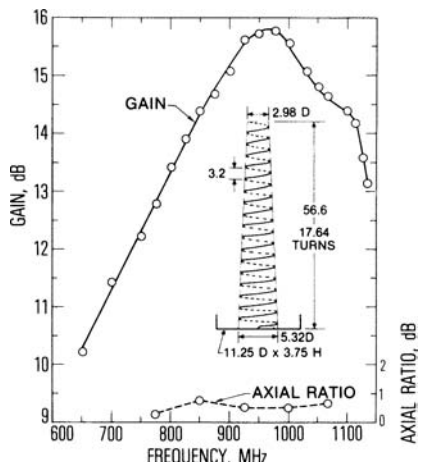


FIGURE 12-22 Gain and axial-ratio characteristics of a 17.64-turn conical helix

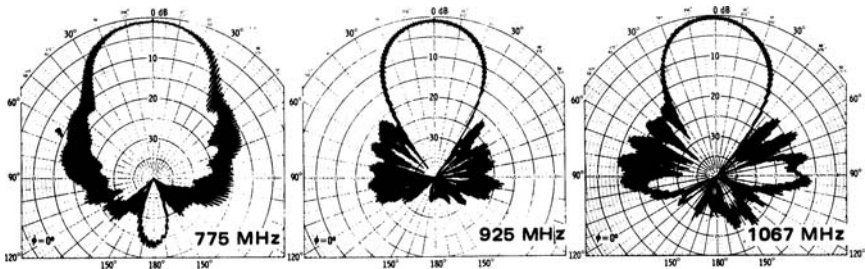


FIGURE 12-23 Radiation patterns of a 17.64-turn conical helix

Two nonuniform helical configurations and the corresponding gain and axial-ratio characteristics are illustrated in Figure 12-24. Configuration *a* consists principally of two uniform-diameter sections [5.28 and 4.13 in (134.1 and 104.9 mm)] joined together by a short tapered transition. This helix may be described as 7-turn (5.28 D) + 2-turn taper (5.28 D to 4.13 D) + 6.64-turn (4.13 D) to 2.98 D). A constant-pitch spacing of 3.2 in (81.28 mm) was maintained in all four helical sections. The performance of this nonuniform helix configuration was optimized over the low-frequency region. The gain is 14.7 ± 0.4 dB from 773 to 900 MHz and remains relatively flat (14.05 ± 0.25 dB) from 900 to 1067 MHz. Note that the gain is constant within ± 1 dB over a frequency ratio of 1.55:1 (1100/710 MHz) as compared with 1.26:1 for a uniform helix. Table 12-1 provides a comparison of the ± 1 -dB bandwidth for the various axial-mode helical antennas. The axial ratio is < 1 dB. The beam-shape and sidelobe characteristics are improved over those of a uniform helix, as illustrated in the patterns of Figure 12-25. Note that the high-frequency cut-off is limited not by the larger 5.28-in-diameter helical section ($C/\lambda = 1.55$ at 1100 MHz) but rather by the smaller 4.12-in-diameter helical section ($C/\lambda = 1.21$ at 1100 MHz). The HPBW is relatively constant, $33^\circ \pm 3^\circ$ over the 773- to 1067-MHz test frequency range.¹²

Configuration *b* of Figure 12-24 consists of a uniform section (5.28-in diameter) plus a tapered section from 5.28- to 2.98-in diameter. The ± 1.1 -dB gain bandwidth is wider than that of configuration *a*, but the gain at the high-frequency end is lower.

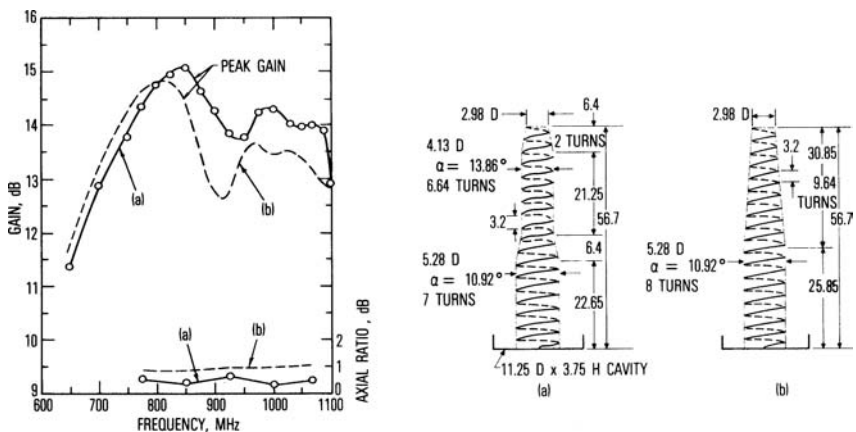


FIGURE 12-24 Gain and axial-ratio characteristics of two 17.64-turn nonuniform-diameter helices

TABLE 12-1

Type of Helix	Frequency Range with ± 1 -dB Gain Variation, MHz	Frequency Ratio, f_{\max}/f_{\min}
Uniform	770–970	1.26:1
Tapered-end	770–980	1.27:1
Continuous-taper	820–1120	1.37:1
Nonuniform	710–1100	1.55:1

12.3 ARRAY OF HELICAL ANTENNAS

A single axial-mode helical antenna produces moderate gain in the end-fire direction (see Figure 12-17). When higher gain is desired, a linear or planar array of helices may be used to produce a fan-shaped beam or a pencil beam, respectively. The array gain depends on the element spacing with respect to the operating wavelength and the element gain. To realize the advantage of the increased gain, the helical elements should be spaced so that the area occupied by each helix is approximately equal to its effective area. To the first-order estimate, this spacing is $\sim \sqrt{G_e / 4\pi\lambda}$, where G_e is the power gain of the individual helices. As the spacing is increased to $> \sqrt{G_e / 4\pi\lambda}$, the array gain approaches the asymptotic value NG_e , where N is the number of elements.³² Some measured data on this type of array have been reported by Harris,^{3,33} and the construction and the radiation characteristics of a typical 2×2 planar array of six-turn helices ($D = 0.3\lambda$ and $\alpha = 14^\circ$) spaced 1.5λ are shown in Figures 12-26 and 12-27, respectively. Also, shaped-beam and steerable arrays employing axial-mode helices have been designed for spacecraft applications.^{22,34,35}

Generally, arrays employing high-gain elements are not suitable for wide-angle scanning because of grating lobes. However, when antenna gain and side-lobes (or grating lobes) are not important, such arrays can be phased to scan over a limited angular range.

For certain applications, such as FM and TV broadcasting, a moderately high-gain, azimuthally omnidirectional pattern is required. A vertical array of side-fire helices mounted concentrically on a conducting cylinder may be employed to produce a narrow-beam pattern in the elevation plane. The helix elements are generally excited in the second- or higher-order modes that produce maximum radiation broadside (normal) to the helix axis and a null along the helix axis. The helix diameter and pitch are chosen so that the length of one turn is equal to an integral number of wavelengths.

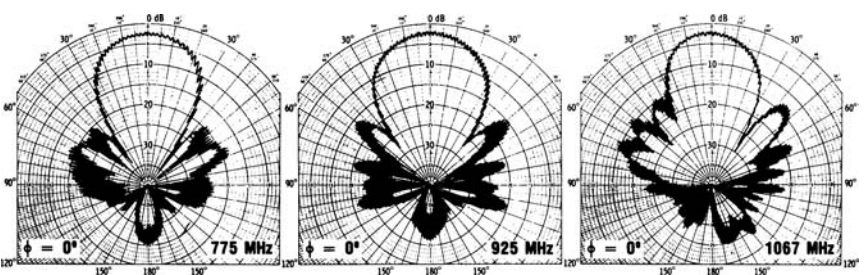


FIGURE 12-25 Radiation patterns of a 17.64-turn nonuniform-diameter helix—configuration *a* of Figure 12-24

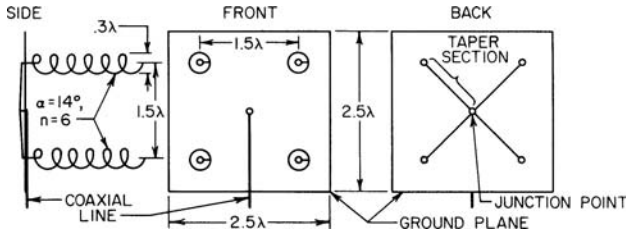


FIGURE 12-26 2×2 array of six-turn helices

If horizontal polarization is desired, the radiating element may be constructed with two oppositely wound helices (one right-handed and the other left-handed), each consisting of a few turns, placed end to end and fed in the center. Design data on this type of vertical array are available in the literature.^{3,36,37}

Vertical helical-array antennas designed to produce omnidirectional, circularly polarized radiation in the horizontal plane for VHF-UHF TV broadcasting have been described by DuHamel.³⁸ In this case, the radiating elements are end-fed, four-turn bifilar helices operating in various modes that are selected so that the helix circumference is about $\lambda/2$ greater than the circumference of the central support cylinder. Depending on the operating frequency, horizontal circularity of ± 0.5 dB to ± 1 dB and an axial ratio of ~ 1.5 dB have been reported.

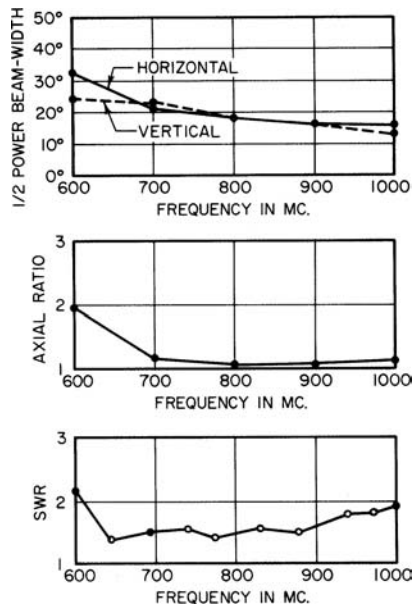


FIGURE 12-27 Measured characteristics of a 2×2 array of six-turn helices

12.4 NORMAL-MODE HELICAL ANTENNAS

When the dimensions of the helix are small compared with the wavelength, the maximum radiation is in a direction normal to the helix axis. Kraus² has shown that the radiation from a short-axial-length helix can be calculated by assuming that the helix is composed of small loops of diameter D and short dipoles of length S . The far field of a short dipole has only an E_θ component,

$$E_\theta = \frac{jBe^{-jkr}}{r} \left(\frac{S}{\lambda} \right) \sin \theta \quad (12-12)$$

and the far field of a small loop has only an E_ϕ component,

$$E_\phi = \frac{Be^{-jkr}}{2r} \left(\frac{\pi D}{\lambda} \right)^2 \sin \theta \quad (12-13)$$

where r is the distance and B is a constant. The E_θ and E_ϕ components are 90° out of phase, and the far field of a small helix is, in general, elliptically polarized. The axial ratio is given by

$$\text{AR} = \frac{|E_\theta|}{|E_\phi|} = \left(\frac{2S}{\lambda} \right) / \left(\frac{\pi D}{\lambda} \right)^2 \quad (12-14)$$

When $\pi D = \sqrt{2S\lambda}$, the axial ratio becomes unity and the radiation is circularly polarized.³⁹ On the other hand, the polarization of the radiated field will be predominantly horizontal ($\text{AR} \rightarrow 0$) when $\pi D \gg \sqrt{2S\lambda}$ or predominantly vertical ($\text{AR} \rightarrow \infty$) when $\pi D \ll \sqrt{2S\lambda}$.

For a normal-mode helix whose dimensions are small compared with wavelength, the current distribution along the helix is approximately sinusoidal.² The terminal impedance is very sensitive to changes in frequency, and the bandwidth is narrow. Nevertheless, a normal helix has been used effectively to reduce the length of thin-wire-type (whip) antennas for personal radio and mobile communications systems in the HF and VHF bands.⁴⁰⁻⁴³ Also, balanced-fed dipole antennas can be constructed by using short-axial-length normal-mode helices when a reduced dipole length is desired.⁴⁴

When a short normal-mode helix is used in conjunction with a ground plane, the polarization is predominantly vertical and the radiation pattern is similar to that of a monopole. The radiation resistance of a short resonant helix above a perfect ground is approximately given⁴⁰ by $(25.3 h/\lambda)^2$, where h is the axial length or height above the ground plane.

Typical feeding arrangements for a helical monopole are shown in Figure 12-28. In the series-fed arrangement, the helix is connected directly to the coaxial input, and an impedance transformer or matching network may be required. In the shunt-fed arrangement, the helix provides a self-matching network by tapping a small portion of the helix. A feed arrangement that employs a bifilar helix to increase the input impedance is described in Hansen,⁴¹ and another design that utilizes a short helical monopole and top-loading wire to produce a self-resonant antenna over the 2–30-MHz frequency range without the need of a matching device is described in Eovine.⁴³

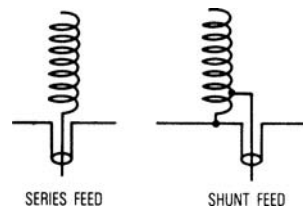


FIGURE 12-28 Typical feed arrangements for a helical monopole

12.5 OTHER HELICAL ANTENNAS

Fractional-Turn Resonant Quadrifilar Helices A class of resonant quadrifilar helices, also referred to as *volute*s, capable of radiating circular polarization with a cardioid-shaped pattern has been described by Kilgus.⁴⁵⁻⁴⁷ The antenna consists of two orthogonal fractional-turn (one-fourth to one turn) bifilar helices excited in phase quadrature. Each bifilar helix is balun-fed at the top, and the helical arms are wires or metallic strips of resonant length ($\ell = m\lambda/4$, $m = 1, 2, 3, \dots$) wound on a small diameter with a large pitch angle. The ends of the helices are open-circuited at the base when $m = \text{odd}$ and short-circuited when $m = \text{even}$. When properly excited, the volute produces a very broad beamwidth with relatively low backlobes and good axial-ratio characteristics over a wide angular range. Since the volute is a resonant structure, the impedance bandwidth is narrow. Typically, the VSWR is $< 2:1$ over a 3 to 5 percent bandwidth. Generally a wider bandwidth can be achieved with larger-diameter wires for a given helix design. This type of antenna has been designed for various applications requiring a wide beamwidth over a relatively narrow frequency range.^{48,49}

Because of its low-backlobe characteristics, a volute can be operated as an isolated antenna or with a ground plane. A half-turn $\lambda/2$ -volute antenna and a typical measured pattern obtained with a rotating linearly polarized source at 488 MHz are shown in Figure 12-29.

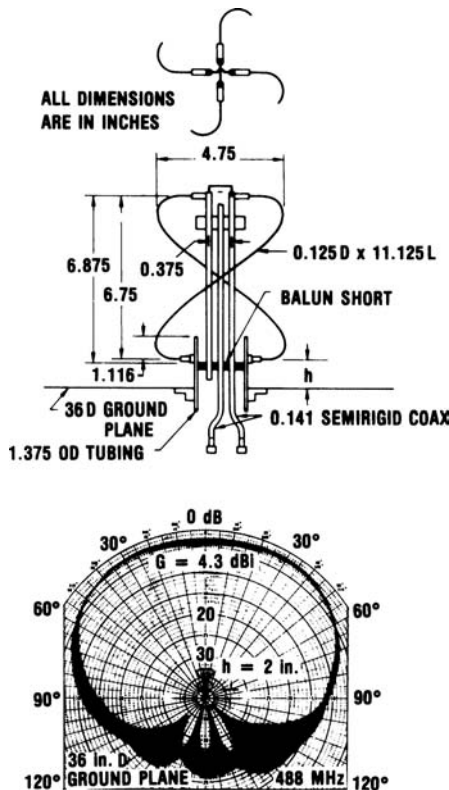


FIGURE 12-29 Half-turn, half-wavelength volute antenna and measured pattern at 488 MHz

The pattern was measured with a 90° hybrid connected to the coaxial input ports of the two bifilar helices. The half-turn, half-wavelength volute is of particular interest because the input impedance of each bifilar can be matched to a 50-Ω coaxial input simply by minor adjustment of the helical-arm lengths without the need of a transformer. The measured VSWR is < 2:1 over ~5 percent bandwidth centered at 488 MHz.⁵⁰

Short Axial-Mode Helices For certain applications requiring a wideband and broadbeam antenna with a relatively small physical size, such as reflector feeds, a short axial-mode helix (two to four turns) housed in a conical or circular cavity similar to that of Figure 12-3 can be employed. As discussed in the subsection “Nonuniform-Diameter and Tapered Helices,” the end of the helix may be slightly tapered to improve the VSWR and axial-ratio performance. A similar cavity-backed short helix designed for flush-mounting applications is described by Bystrom and Berntsen.⁵¹ A dual-helix feed system that utilizes two concentric helices to provide operation over two narrow-frequency bands is described by Holland.⁵²

A helical feed design for a paraboloidal reflector that utilizes a four-turn backfire bifilar helix is described in King and Wong.⁵³ This feed was designed for a spacecraft antenna operating in the 240- to 270-MHz band.* An attractive feature of the backfire bifilar-helix feed is that it does not require a cavity or ground plane, thus minimizing blockage effects and improving reflector-antenna efficiency. Based on the measured gain results reported in King and Wong,⁵³ antenna efficiency of 64 to 70 percent is achievable over a 12 percent bandwidth. Similar reflector feed designs employing a backfire monofilar helix have been described by Johnson and Cotton⁵⁴ and Nakano et al.⁵⁵

Short-axial-mode helices with less than three turns are useful for array applications, particularly when circular polarization is desired. Nakano et al.⁵⁶ have demonstrated that a two-turn helix with a pitch angle as small as 4° can provide an axial ratio of less than 3 dB over a 12 percent bandwidth. A combination of low pitch and a small number of turns results in a low-profile configuration. A circular array using the two-turn helix as the radiating element is described in Nakano et al.⁵⁷

The performance of a short-axial-mode helix is generally limited by its axial ratio and bandwidth characteristics. The helix can be designed to operate over a larger bandwidth if a larger pitch angle is employed and with some compromise on the minimum achievable axial ratio. Figure 12-30 depicts the typical measured radiation patterns of a 1.59-turn helix with a pitch angle of 11° on a 30- by 30-in (76.2- by 76.2-cm) ground plane.† The axial ratio varies from 1 to 3 dB in the 1.8- to 2.4-GHz band, which represents a bandwidth of nearly 30 percent. The helix was constructed using 0.081-in (2.06-mm) diameter copper wire wound on a Styrofoam cylinder. Additional laboratory experiments indicated that the axial ratio characteristics of short-axial-length helices can be improved by loading the helix end with a small taper, a small loop, or a flat spiral.

As mentioned previously, when a short-axial-mode helix is used as a single widebeam element, it can be placed in a circular cavity instead of a ground plane. Limited laboratory measurements showed that there are no significant changes in axial ratio characteristics. Also, when a short-axial-mode helix with an imperfect axial ratio is used as the radiating element for an array, the overall axial ratio characteristics of the array can be improved by

*By TRW Systems Group, Redondo Beach, Calif.

†H. E. King and J. L. Wong, Aerospace Corporation sponsored research results.

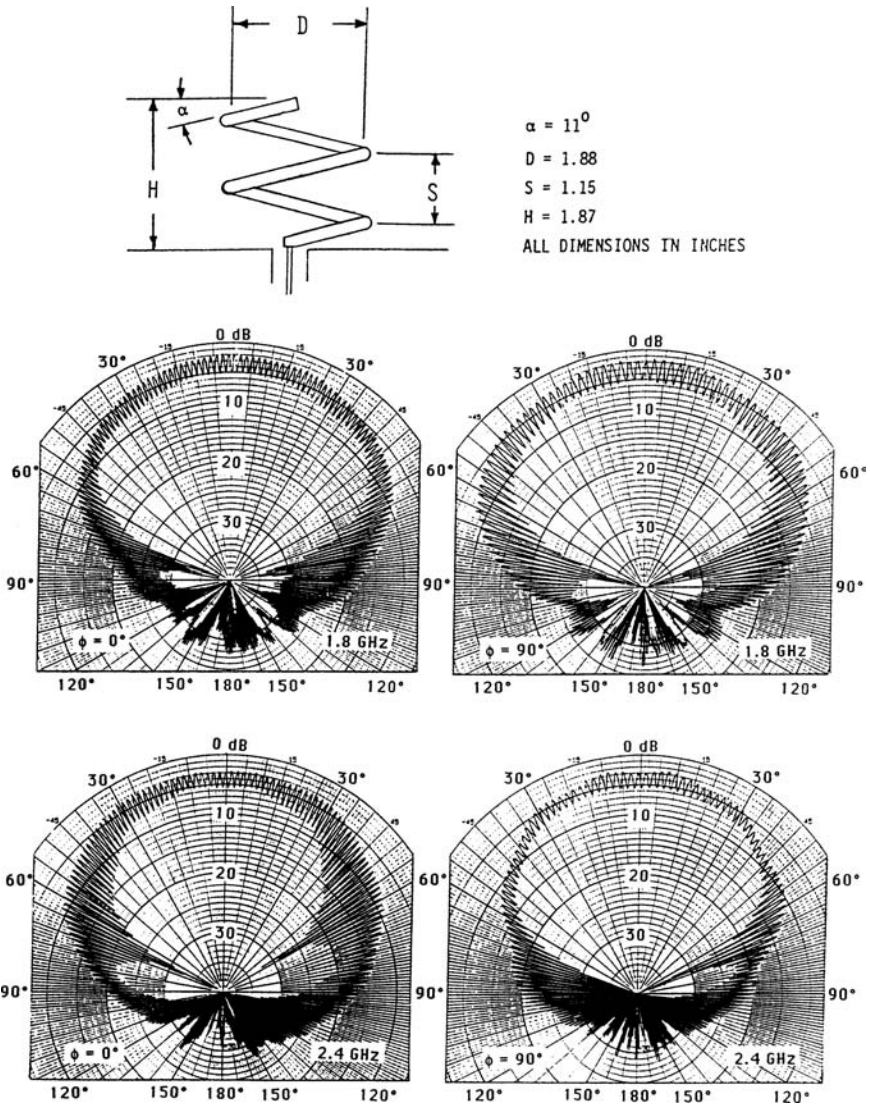


FIGURE 12-30 Radiation patterns of a 1.59-turn helix on a 30- by 30-in ground plane

appropriately arranging the relative orientation (roll position) and adjusting the excitation phase of the individual elements within the array.⁵⁸

Helicone Antennas An axial-mode helix can be operated in conjunction with a conical horn to provide a broadband antenna with low sidelobes.⁵⁹ The axial length of the helix and the horn are approximately the same. Sidelobe levels of the order of -25 dB have been reported. Design and performance data are available in the literature.⁶⁰

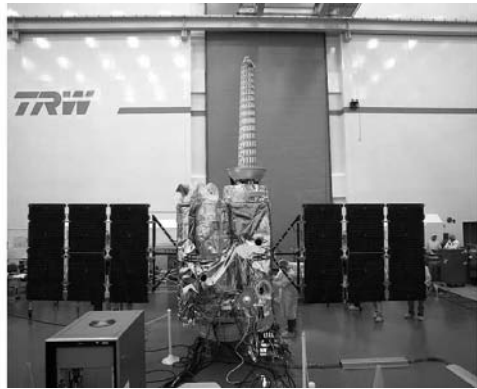


FIGURE 12-31 SKYNET 4 UK MOD spacecraft UHF helical antenna designed and built in 1984 by Sultan et al^{61–63} at CAL (Canadian Astronautics Limited—now EMS Technologies Ottawa). Similar antennas were later deployed on NATO and The National Reconnaissance Office. GeoLITE satellite built by TRW/NRO, <http://spaceflightnow.com/delta/d285/010517geolite.html>. Posted May 17, 2001.

12.6 SATELLITE WIDEBAND HELICAL ANTENNA DESIGN

For space applications the optimization of mass, size, and gain of the antenna are essential. In addition, spacecraft such as SKYNET, NATO, and GeoLITE require, at launch, a stowed tapered helical antenna that is deployable on-orbit, by extending a center rod and using a metal tape as radiating elements. Canadian Astronautics Ltd.^{61–63} built and tested a wide range of single and multisection, single and multipitch uniform cylindrical and tapered helices. To achieve wide bandwidth, stepped cylindrical⁶¹ instead of tapered helical antennas were proposed with multisections and multipitch angles. One of the resulting configurations was a pair of concentric helices sharing the same ground reflector. This 2-wavelength-long antenna provides 12-dBi gain at 760 and 1290 MHz, with 16 dB isolation between the frequencies to reduce passive intermodulation.

To date, there are no simple mathematical solutions for designing helical antennas. Unfortunately, a 1-dBi uncertainty can lead to as much as one-third of the helix length. Most often, modeling software based on NEC⁶⁴ is used to compute the currents on the helix wire or metallic tapes forming the helix.

12.7 RECENT DEVELOPMENTS

Cardoso and Safaai-Jazi^{65–67} found that winding a helix on a spherical, rather than a cylindrical, surface can produce designs that maintain good circular polarization over a broader beamwidth. Hemispherical designs have also been proposed and have the advantage of reduced height and operation over a wide angular region.^{68–72} Various methods have been presented for reducing the size^{71–73} and improving the bandwidth of helical antennas,^{74–78} including the use of wideband impedance matching.^{79–81} Finally, it is worthy of note that in 1995, Kraus presented an axial mode helix design that did not require a ground plane.⁸²

REFERENCES

1. J. D. Kraus, "Helical Beam Antennas," *Electronics*, vol. 20 (April 1947): 109–111.
2. J. D. Kraus, "The Helical Antenna," *Antennas*, Chap. 7 (New York: McGraw-Hill, 1950).
3. E. F. Harris, "Helical Antennas," *Antenna Engineering Handbook*, 1st Ed., H. Jasik (ed.), Chap. 7 (New York: McGraw-Hill, 1961).
4. A. G. Holtum, Jr., "Improving the Helical Beam Antenna," *Electronics* (April 29, 1960).
5. A. T. Adams, R. K. Greenough, R. F. Wallenberg, A. Mendelovicz, and C. Lumjiak, "The Quadrifilar Helix Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (March 1974): 173–178.
6. C. C. Kilgus, "Shaped-Conical Radiation Pattern Performance of the Backfire Quadrifilar Helix," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (May 1975): 392–397.
7. C. W. Gerst and R. A. Worden, "Helix Antennas Take Turn for Better," *Electronics* (August 22, 1966).
8. W. T. Patton, "The Backfire Bifilar Helical Antenna," Ph.D. dissertation, University of Illinois, Urbana, October 1963.
9. J. Y. Wong and S. C. Loh, "Radiation Field of an Elliptical Helical Antenna," *IRE Trans. Antennas Propagat.*, vol. AP-7 (January 1959): 46–52.
10. B. A. Munk and L. Peters, "A Helical Launcher for the Helical Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-16 (May 1968): 362–363.
11. H. E. King and J. L. Wong, "Characteristics of 5- to 35-Turn Uniform Helical Antennas," Aerospace Corp. Tech. Rep. TR-0078(3724-01)-2, DDC AD A046487 (June 1977); or "Characteristics of 1 to 8 Wavelength Uniform Helical Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-28 (March 1980): 291–296.
12. J. L. Wong and H. E. King, "Broadband Quasi-Taper Helical Antennas," Aerospace Corp. Tech. Rep. TR-0077(2724-01)-2, DDC AD A046067 (September 30, 1977); and *IEEE Trans. Antennas Propagat.*, vol. AP-27 (January 1979): 72–78.
13. O. J. Glasser and J. D. Kraus, "Measured Impedances of Helical Antennas," *J. App. Phys.*, vol. 19 (February 1948): 193–197.
14. J. A. Marsh, "Measured Current Distributions on Helical Antennas," *IRE Proc.*, vol. 39 (June 1951): 668–675.
15. J. D. Kraus, "A 50-Ohm Input Impedance for Helical Beam Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (November 1977): 913–914.
16. H. Nakano, J. Yamauchi, and H. Mimaki, "Tapered Balanced Helices Radiating in the Axial Mode," *Dig. Int. Symp. Antennas Propagat.* (1980): 700–703.
17. D. J. Angelakos and D. Kajfez, "Modifications on the Axial-Mode Helical Antenna," *IEEE Proc.*, vol. 55 (April 1967): 558–559.
18. W. W. Hansen and J. R. Woodyard, "A New Principle in Directional Antenna Design," *IRE Proc.*, vol. 26 (March 1938): 333–345.
19. R. J. Stegen, "The Gain-Beamwidth Product of an Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-12 (July 1964): 505–506.
20. T. S. M. Maclean and R. G. Kouyoumjian, "The Bandwidth of Helical Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-7, special supplement (December 1959): S379–S386.
21. G. C. Jones, "An Experimental Design Study of Some S- and X-Band Helical Aerial Systems," *IEE Proc.*, vol. 103, part B (November 1956): 764–771.
22. K. G. Schroeder and K. H. Herring, "High Efficiency Spacecraft Phase Array," AIAA Pap. 70-425, AIAA 3d Comm. Satellite Syst. Conf., Apr. 6–8, 1970.
23. T. S. M. Maclean, "An Engineering Study of the Helical Aerial," *IEE Proc.*, vol. 110 (January 1963): 112–116.
24. S. Sander and D. K. Chang, "Phase Center of Helical Beam Antennas," *IRE Conv. Rec.*, part I (1958): 152–157.

25. C. Donn, "A New Helical Antenna Design for Better On- and Off-Boresight Axial Ratio Performance," *IEEE Trans. Antennas Propagat.*, vol. AP-28 (March 1980): 264–267.
26. J. L. Wong and H. E. King, "Empirical Helix Antenna Design," *Dig. Int. Symp. Antennas Propagat.* (1982): 366–369.
27. J. S. Chatterjee, "Radiation Field of a Conical Helix," *J. App. Phys.*, vol. 24 (May 1953): 550–559.
28. H. S. Barsky, "Broadband Conical Helix Antennas," *IRE Conv. Rec.*, part I (1959): 138–146.
29. K. F. Lee, P. F. Wong, and K. F. Larm, "Theory of the Frequency Response of Uniform and Quasi-Taper Helical Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-30 (September 1982): 1017–1021.
30. J. M. Tranquilla and G. B. Graham, "Swept-Frequency Radiation Pattern Anomalies on Helical Antennas," *Can. Elect. Eng. J.*, 1982.
31. J. L. Wong and H. E. King, "Broadband Helical Antennas," (September 25, 1979): U.S. Patent 4,169,267.
32. H. E. King and J. L. Wong, "Directivity of a Uniformly Excited $N \times N$ Array of Directive Elements," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (May 1975): 401–403.
33. E. F. Harris, "Helical Beam Antenna Performance," *Commun. Eng.* (July–August 1953): 19–20, 44–45.
34. C. T. Brumbaugh, A. W. Love, G. M. Randall, D. K. Waineo, and S. H. Wong, "Shaped Beam Antenna for the Global Positioning Satellite System," *Dig. Int. Symp. Antennas Propagat.* (1976): 117–120.
35. C. Donn, W. A. Imbriale, and G. Wong, "AS-Band Phased Array Design for Satellite Applications," *Dig. Int. Symp. Antennas Propagat.* (1977): 60–63.
36. L. O. Krause, "Sidefire Helix UHF TV Transmitting Antenna," *Electronics* (August 1951): 107.
37. H. G. Smith, "High-Gain Side Firing Helical Antennas," *AIEE Trans.*, part I, *Commun. & Electronics*, vol. 73 (May 1954): 135–138.
38. R. H. DuHamel, "Circularly Polarized Helix and Spiral Antennas," (September 16, 1975): U.S. Patent 3,906,506.
39. H. A. Wheeler, "A Helical Antenna for Circular Polarization," *IRE Proc.*, vol. 35 (December 1947): 1484–1488.
40. A. G. Kandoian and W. Sichak, "Wide Frequency Range Tuned Helical Antennas and Circuits," *IRE Conv. Rec.*, part 2 (1953): 42–47.
41. L. H. Hansen, "A New Helical Ground Plane Antenna for 30 to 50 MHz," *IRE Trans. Veh. Commun.*, vol. VC-10 (August 1961): 36–39.
42. D. A. Tong, "The Normal-Mode Helical Aerial," *Radio Commun.* (July 1974): 432–437.
43. M. Eovine, "Helical-Monopole HF Antenna," Dept. of Army Proj. 3F37-01-001-04, Chu Associates (October 15, 1962).
44. Y. Hiroi and K. Fujimoto, "Practical Usefulness of Normal Mode Helical Antenna," *Dig. Int. Symp. Antennas Propagat.* (1976): 238–241.
45. C. C. Kilgus, "Multielement, Fractional Turn Helices," *IEEE Trans. Antennas Propagat.*, vol. AP-16 (July 1968): 499–501.
46. C. C. Kilgus, "Resonant Quadrifilar Helix," *IEEE Trans. Antennas Propagat.*, vol. AP-17 (May 1969): 349–351.
47. C. C. Kilgus, "Resonant Quadrifilar Helix Design," *Microwave J.*, vol. 13 (December 1970): 49–54.
48. C. C. Kilgus, "Spacecraft and Ground Station Applications of the Resonant Quadrifilar Helix," *Dig. Int. Symp. Antennas Propagat.* (1974): 75–77.
49. R. W. Bricker and H. H. Rickert, "A S-Band Resonant Quadrifilar Antenna for Satellite Communication," *Dig. Int. Symp. Antennas Propagat.* (1974): 78–82.
50. J. L. Wong, H. E. King, C. E. Ermatinger, et al, "A Jam-Resistant UHF Antenna for the GPS Integrated Transfer Subassembly," Aerospace Corp. Tech. Rep. TOR-0082 (2414-02)-1 (April 1982).

51. A. Bystrom, Jr., and D. G. Berntsen, "An Experimental Investigation of Cavity-Mounted Helical Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-4 (January 1956): 53–58.
52. J. Holland, "Multiple Feed Antenna Covers L, S, and C Band Segments," *Microwave J.*, vol. 24 (October 1981): 82–85.
53. H. E. King and J. L. Wong, "240–400 MHz Antenna System for the FleetSatCom Satellites," *Dig. Int. Symp. Antennas Propagat.* (1977): 349–352.
54. R. C. Johnson and R. B. Cotton, "A Backfire Helical Feed," *IEEE Trans. Antennas Propagat.*, vol. AP-32 (October 1984): 1126–1128.
55. H. Nakano, J. Yamauch, and H. Mimaki, "Backfire Radiation from a Monofilar Helix with a Small Ground Plane," *IEEE Trans. Antennas Propagat.*, vol. AP-36 (October 1988): 1359–1364.
56. H. Nakano et al., "Extremely Low-Profile Helix Radiating a Circularly Polarized Wave," *IEEE Trans. Antennas Propagat.*, vol. AP-39 (June 1991): 754–757.
57. H. Nakano et al., "An Extremely Low-Profile Helical Array Antenna," *Dig. IEEE Int. Symp. Antennas Propagat.* (1990): 702–705.
58. C. E. Ermatinger and W. G. Scott, "The Crossed Ellipse Array—A System for Reduction of Polarization Axial Ratio," *Dig. IEEE Int. Symp. Antennas Propagat.* (1972): 21–24.
59. K. R. Carver, "The Helicone—A Circularly Polarized Antenna with Low Sidelobe Level," *IEEE Proc.*, vol. 55, no. 4 (April 1967): 559.
60. K. R. Corver and B. M. Potts, "Some Characteristics of the Helicone Antenna," *Dig. Int. Symp. Antennas Propagat.* (1970): 142–150.
61. N. Sultan, M. H. Moody, J. Whelpton, and C. B. Hodgson, "Design of Concentric Helical Antennas for Maximum Gain and Isolation Over Wide Frequency Separation," NTC '83, IEEE AES National Telesystems Conference, San Francisco (November 1983): 12–17.
62. N. Sultan, M. H. Moody, J. Whelpton, and C. B. Hodgson, "Novel Broadband Double Pitch Cylindrical Helical Antenna for Satellite & Ground Applications," 1984 International IEEE Antennas and Propagation Symposium AP-S and National Radio Science Meeting, Boston, June 1984.
63. N. Sultan, J. Whelpton, and M. H. Moody, "Design of Broadband Tapered Helical Antennas," IEE Fourth International Conference on Antennas and Propagation - ICAP 85, Coventry, U.K., April 1985.
64. C. W. Trueman, N. Sultan, S. J. Kubina, and T. Pellerin, "Software for Modeling Helix Antennas with NEC and Validation by Measurements," 12th Annual Review of Progress in Applied Computational Electromagnetics, The Applied Computational Electromagnetics Society, ACES, Monterey, Calif., March 18, 1996.
65. J. C. Cardoso and A. Safaai-Jazi, "Spherical Helical Antenna with Circular Polarization over a Broad Beam," *Electronics Letters*, vol. 29 (February 1993): 325–326.
66. A. Safaai-Jazi and J. C. Cardoso, "Radiation Characteristics of a Spherical Helical Antenna," *IEE Proc. Microwaves, Ant. and Prop.*, vol. 143 (February 1996): 7–12.
67. E. Weeratumanoon and A. Safaai-Jazi, "Truncated Spherical Helical Antennas," *Electronic Letters*, vol. 36 (March 2000): 607–609.
68. H. T. Hui, K. Y. Chan, and E. K. N. Yung, "The Low-Profile Hemispherical Antenna with Circular Polarization Over a Wide Angle Range," *IEEE Trans. Antennas Propagat.*, vol. AP-51 (June 2003): 1415–1418.
69. H. T. Hui, E. K. N. Yung, C. L. Law, Y. S. Koh, and W. L. Koh, "Design of a Small and Low-Profile 2×2 Hemispherical Helical Antenna Array for Mobile Satellite Communications," *IEEE Trans. Antennas Propagat.*, vol. AP-52 (January 2004): 346–348.
70. Z. Yingbo and H. T. Hui, "A Printed Hemispherical Helical Antenna for GPS Receivers," *IEEE Microwave and Wireless Components Letters*, vol. 15 (January 2005): 10–12.
71. J. C. Louvigne, J. C. Sharaiha, and D. Thouroude, "Broadband Compact Printed Quadrifilar Helical Antenna for Balloon Campaign Application," *Electronics Letters*, vol. 38 (August 2002): 944–945.
72. D. K. C. Chew and S. R. Saunders, "Meander Line Technique for Size Reduction of Quadrifilar Helix Antenna," *Antennas and Wireless Propagation Letters*, vol. 1 (2002): 109–111.

73. M. Amin and R. Cahill, "Side-Fed Bifilar Helix Antenna," *IEEE Microwave and Wireless Components Letters*, vol. 15 (December 2005): 912–915.
74. J. C. Louvigne and A. Sharaiha, "Broadband Tapered Printed Quadrifilar Helical Antenna," *Electronics Letters*, vol. 37 (July 2001): 932–933.
75. J. A. Dobbins and R. L. Rogers, "Folded Conical Helical Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-49 (December 2001): 1777–1781.
76. S. H. Sim, C. Y. Kang, S. J. Yoon, Y. J. Yoon, and H. J. Kim, "Broadband Multilayered Ceramic Chip Antenna for Handsets," *Electronics Letters*, vol. 38 (February 2002): 205–207.
77. S. D. Rogers and C. M. Butler, "Wide-Band Sleeve-Cage and Sleeve-Helical Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-50 (October 2002): 1409–1414.
78. K. Noguchi, S. I. Betsudan, T. Katagi, and M. Mizusawa, "A Compact Broad-Band Helical Antenna with Two-Wire Helix," *IEEE Trans. Antennas Propagat.*, vol. AP-51 (September 2003): 2176–2181.
79. K. Wong and W. Chen, "Study of an Internally Matched Helical Beam Antenna," *IEEE Trans. Antennas and Propagat.*, vol. AP-39 (June 1991): 811–814.
80. M. Notter and K. M. Keen, "Impedance Matching Arrangement for Quadrifilar Helix Antennas," *Electronics Letters*, vol. 38 (April 2002): 353–354.
81. M. Hosseini, M. Hakkak, and P. Rezaei, "Design of a Dual Band Quadrifilar Helix Antenna," *Antennas and Wireless Propagation Letters*, vol. 4 (2005): 39–42.
82. J. D. Kraus, "A Helical-Beam Antenna Without a Ground Plane," *IEEE Antennas and Propagation Magazine*, vol. 37 (April 1995): 45.

Chapter 13

Frequency
Independent
Antennas

Dejan S. Filipovic

University of Colorado at Boulder

Tom Cencich

Lockheed Martin Space Systems Company

CONTENTS

13.1	INTRODUCTION	13-2
13.2	SPIRAL ANTENNAS	13-7
13.3	LOG-PERIODIC ANTENNAS	13-35
13.4	DUAL-CIRCULARLY POLARIZED ANTENNAS	13-52

13.1 INTRODUCTION

Theoretically, frequency independent (FI) antennas have near-field and far-field characteristics independent of frequency. Thus, an FI antenna must have an infinitely large aperture to eliminate the low-frequency limit. It must also have a very fine and infinitely small feed region to remove the high-frequency limit. To eliminate length-dependent scaling, an FI antenna is exclusively described by angles. To reduce and ideally eliminate the far-field contamination, the residual currents must decay to zero after passing through the active (principal or main) region and before entering the next-higher order radiating region. The complete radiation from a single active region, also known as constructive radiation, is however not possible. Finally, an ideal FI antenna must have a frequency independent feed.

Certainly, no practically realizable antenna can fulfill these requirements. Thus, realistic FI antennas are radiators having virtually invariant impedance and pattern characteristics over very wide instantaneous bandwidths or consecutive multiple logarithmically periodic bands. Consistent impedance and pattern characteristics over 100:1 bandwidths have been successfully demonstrated, so that the system bandwidth is commonly limited by the electronics, rather than by the antenna. Owing to the previously mentioned properties, FI antennas have become irreplaceable components of many, predominantly military, systems. This chapter outlines the fundamental principles of frequency independence and reviews many antenna structures belonging to this class.

Basic Principles

The following principles underline FI antennas and their performance¹⁻⁷:

- Geometry scaling
- Geometry description by angles
- Truncation of residual currents
- Self-complementarity

Clearly, these principles disregard effects of the balun feed, which must be frequency independent over the same finite bandwidth as the antenna. In a nondispersive medium, the current distribution on an FI antenna as a function of frequency remains fixed only if both the dimensions of the antenna and the operating wavelength are changed by the same fraction. This ensures that the antenna's electrical dimensions remain invariant to changes in operating frequency. Consequently, the invariance of currents (sources) with normalized frequency leads to near- and far-field invariance with frequency. This scaling concept can be readily extended for inhomogeneous structures composed of metals and dielectrics if all dimensions are scaled inversely proportional to frequency.

To obtain continuous scaling of antenna performance with frequency, the radiating structure of the antenna must be specified only by angles. This way the frequency scaling, which for antennas with a lineal structural form results in resonance scaling, corresponds to rotation and translation of the basic structure. Exact mathematical formulation for frequency independent planar and non-planar antennas, based on the previous principles, will be derived later in this section.

Theoretically, an infinitely broadband FI antenna, defined purely by angles, must be infinitely large with infinitely fine center details. However, for any practical use where the bandwidth of the system electronics and allowed space are finite, the antenna bandwidth may be reduced accordingly. To maintain the FI performance while having reduced/finite size,

the antenna must obey the truncation principle. That is, the residual currents must decay to zero after they have been radiated from the principal active region. By allowing significant existence of residual currents, impedance and pattern degradations can occur. Impedance bandwidth is affected when these currents reach antenna ends and reflect back to the feeder. This effect can be mitigated, as will be shown later. Pattern degradation occurs due to the radiation from the higher order regions when the antenna is large enough and from the residual currents reflected from the arm ends. However, if the residual currents decay to zero or to a very small value, then the antenna aperture can be terminated at the finite distance and reflection from the ends will be eliminated. This way, as the frequency is increased, the radiation pattern of a finite, truncated structure approaches limiting form; that is, the pattern of an infinite structure.

The truncation principle establishes the low-frequency limit of an FI antenna. At the lower end of operation, two important situations must be addressed. First, the high-frequency regions become part of the antenna feed system. More clearly, the low-frequency wave, while traveling to the designated active region, uses the path between the feed and active region as a transmission line. Second, at some frequency the currents will start to interact with the antenna ends before decaying to zero. Below this frequency, the antenna will still radiate, but its current distribution will be disturbed and antenna impedance and pattern will be inconsistent with frequency. To mitigate end effects, a lossy termination is typically integrated with the end of the radiator. The upper-frequency limit is determined by the precision with which the balun and feed region can be fabricated.

Self-complementarity is not a necessary condition for FI performance. This principle states that if the interchange of metallic and nonmetallic parts on a planar antenna embedded in a homogeneous medium leaves the geometry unchanged except for a rotation equal to one half of the angular periodicity, then the respective impedances of two complementary antennas are related by

$$Z_{\text{METAL}} Z_{\text{SLOT}} = \left(\frac{\eta}{2} \right)^2 \quad (13-1)$$

where η is the plane wave impedance for the homogeneous medium (for an antenna in air $\eta = 120\pi \Omega$, and the impedance is $Z_{\text{METAL}} = Z_{\text{SLOT}} = 188.5 \Omega$). Equation 13-1 is based on Booker's extension of Babinet's principle⁶ and has been widely utilized in designing antennas with a broadband impedance match. Self-complementary two-arm spiral, four-arm sinuous, and four-arm planar log-periodic antennas are shown in Figure 13-1.

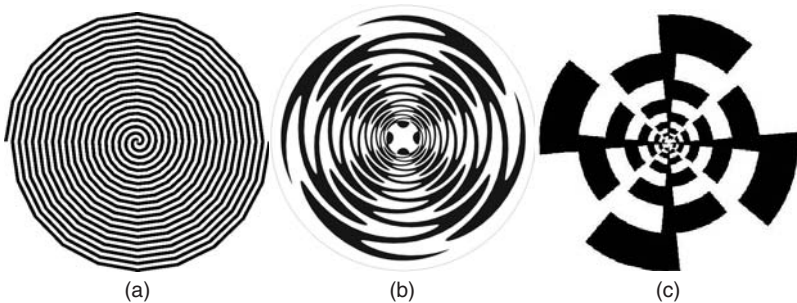


FIGURE 13-1 Examples of self-complementary planar structures: (a) two-arm spiral, (b) four-arm sinuous, (c) and four-arm planar log-periodic

Antennas with minimum impedance and pattern variations over a frequency range repeated with the logarithmic periodicity also adhere to the FI principles. Both infinitely large and infinitesimally fine structure is needed to comply with the theoretical requirements; however, for practical purposes, these conditions are relaxed. The connection between the two apparently different FI conceptual approaches is mathematically well supported. Specifically, the conical screw FI antenna structure allows log-periodic antennas to become FI. Finally, note that this principle is more a rule for evaluation rather than for design, and it implies that a properly designed FI antenna should have identical performance at frequencies related by the (logarithmic) periodicity of the antenna with as little performance variation as possible within each period.

The two arms of a self-complementary six-arm log-periodic structure are depicted in Figure 13-2. The geometric ratio, or growth rate, is denoted by τ and is the ratio of frequencies between which *small* variations in antenna performance occur. The growth rate is also equal to the ratio of the lower (smaller) and the next higher (larger) repeatable structural dimension (i.e., $\tau = f_n / f_{n+1} = r_{n+1} / r_n$). To be able to truncate the log-periodic geometry, the antenna must be designed to eliminate radiation in the direction of the antenna's low-frequency expansion. If the pattern null is not produced in the truncated direction, the currents will also exist in infinity and the end effect will be significant. This violates the truncation principle.

Mathematics of Frequency Independence

The derivation of frequency independence given in this section follows the work from Rumsey,^{1,2} Elliot,⁷ and DuHamel and Isbell⁸ and is valid for spherical coordinate systems, r , θ , and ϕ . The following assumptions have been made:

- The antenna ports are symmetric around the origin and the distance between the ports is infinitesimally small.
- $\theta = 0^\circ$ is the axis of symmetry.
- The antenna is infinite.
- The antenna arms are made of a lossless conductor.
- The medium is homogeneous and isotropic.
- The outline of antenna arms can be described with the parametric formula $r = r(\phi, \theta)$, where r , ϕ , and θ are spherical coordinates and $r(\phi, \theta)$ is the radius vector to any point on the antenna arms.

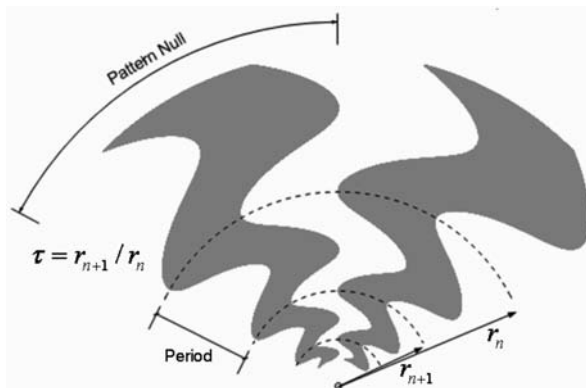


FIGURE 13-2 Schematic of two-arm geometry. The full structure is that of a planar six-arm self-complementary log-periodic saw-tooth antenna.

To scale the antenna performance to α times lower in frequency, i.e. $f' = f/\alpha$, the electrical dimensions of the original must be maintained. Thus, all defining points on the antenna arms must be multiplied with the same factor. The radius vector now becomes

$$r' = \alpha \cdot r(\phi, \theta) \quad (13-2)$$

Figure 13-3 depicts an example of original and scaled geometry with $\alpha = 2$. The solid circle in the middle denotes the infinite precision of the feeding region. Therefore, the scaling is equivalent to the rotation of the geometry by the angle ϕ' around the fixed antenna terminals. Mathematically,

$$r' = \alpha \cdot r(\phi, \theta) = r(\phi + \phi', \theta) \quad (13-3)$$

Equation 13-3 clearly shows that the two antennas will have the same performance at two frequencies related by α , except for the associated azimuth rotation ϕ' . Additionally, the scaling factor α and associated azimuth rotation ϕ' are mutually dependent and invariant to any other parameter of the antenna. Thus, after taking partial derivatives ($\partial/\partial\phi$ and $\partial/\partial\phi'$) of both sides in Eq. 13-3,

$$\alpha \frac{\partial r(\phi, \theta)}{\partial \phi} = \frac{\partial r(\phi + \phi')}{\partial \phi} = \frac{\partial r(\phi + \phi')}{\partial(\phi + \phi')} \frac{\partial(\phi + \phi')}{\partial \phi} = \frac{\partial r(\phi + \phi')}{\partial(\phi + \phi')} \quad (13-4)$$

$$r(\phi, \theta) \frac{d\alpha}{d\phi'} = \frac{\partial r(\phi + \phi')}{\partial \phi'} = \frac{\partial r(\phi + \phi')}{\partial(\phi + \phi')} \frac{\partial(\phi + \phi')}{\partial \phi'} = \frac{\partial r(\phi + \phi')}{\partial(\phi + \phi')} \quad (13-5)$$

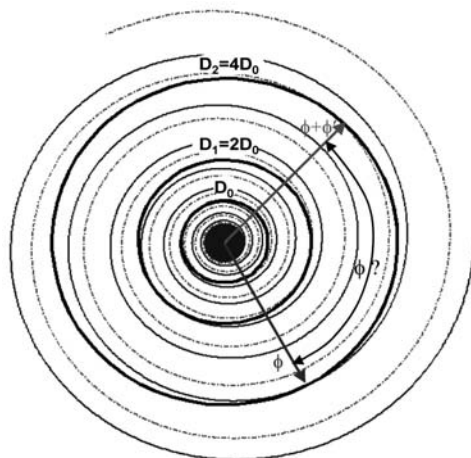


FIGURE 13-3 Notion of scaling through rotation of an infinite FI geometry

Because the right-hand sides of Eq. 13-4 and Eq. 13-5 are the same,

$$\frac{1}{r(\phi, \theta)} \partial r(\phi, \theta) = \left(\frac{1}{\alpha} \frac{d\alpha}{d\phi} \right) \partial \phi = a \partial \phi \quad (13-6)$$

where $a = \alpha^{-1} d\alpha/d\phi$ is a constant. The solution to Eq. 13-6 is

$$r = r(\phi, \theta) = f(\theta) \exp(a\phi) \quad (13-7)$$

where $f(\theta)$ is the arbitrary functional dependence of antenna structure on angle θ . An antenna described by Eq. 13-7 is frequency independent. For the example shown in Figure 13-3, the infinitely large structures represented by solid and dotted lines follow the same growth rate a and the dependence with elevation angle $f(\theta) = \text{constant}$. When the scaling factor is $\alpha = 2$, the original equiangular spiral (solid line) *rotates* in azimuth by $\phi' = \log 2/a$ to conform to the dotted spiral structure. Thus, the antenna scaling is equivalent to a rotation, and the performance at any two frequencies f_1 and αf_1 must be the same except for the rotation ϕ' in azimuth. The rotation of the antenna pattern with frequency is experimentally verified with the equiangular (log-) and other spiral antennas.

By closely examining Eq. 13-7, the following special cases can be derived:

- *Planar equiangular spiral* $f'(\theta) = A\delta(\pi/2 - \theta)$, where A is constant and $\delta(\pi/2 - \theta)$ is the Dirac delta function clearly defining a planar structure.
- *Conical equiangular spiral* $f'(\theta) = A\delta(\beta - \theta)$, where β is the apex angle of the cone while the spiral arms are projected onto the cone surface.
- *Conical screw* The equiangular spiral is projected onto the volume (not surface) between the two concentric cones having different apex angles.
- *Log-periodic planar sheet* Taking $\phi = \text{constant}$ cut through a conical screw with axis on the θ axis. A drawing of the self-complementary log-periodic antenna⁸ is shown in Figure 13-4. Rotation of this planar structure around the arm's symmetry axis will give a

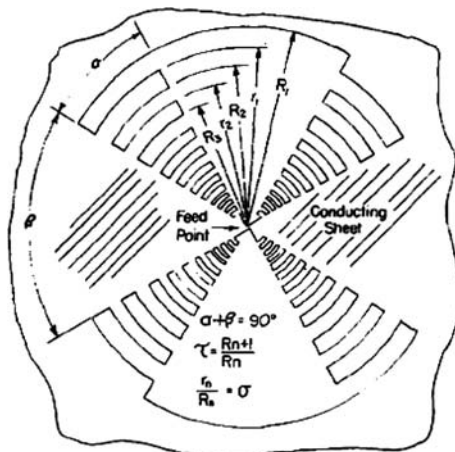


FIGURE 13-4 Drawing of the first planar two-arm sheet log-periodic antenna (after R. H. DuHamel and D. E. Isbell⁸ © IEEE 1957)

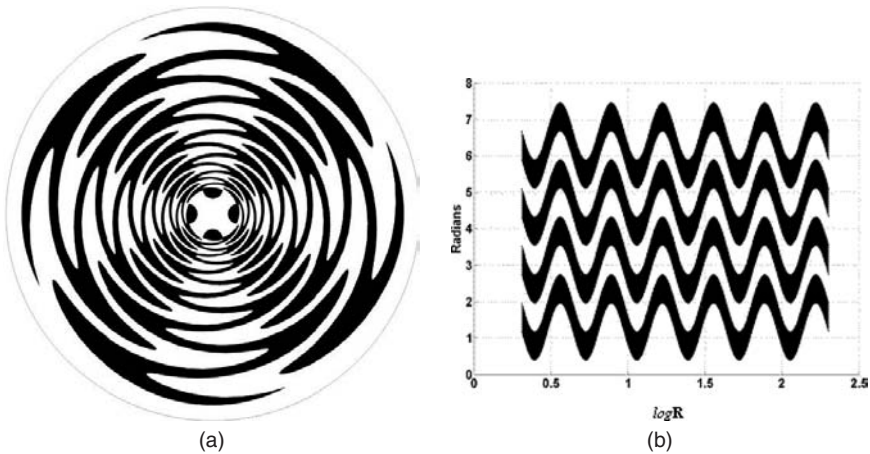


FIGURE 13-5 A four-arm self-complementary sinuous antenna (a) and its \log transformation (b)

biconical screw. Log-periodic structures can be obtained by utilizing the transformation $z = \ln(w)$, where $z(x,y)$ and $w(\rho,\phi)$ represent two spaces. This transformation is depicted in Figure 13-5 for a four-arm self-complementary sinuous antenna.

13.2 SPIRAL ANTENNAS

L to Ku band radar developments after the Second World War created a need for compact, flush-mounted antennas and their use for electronic countermeasures (ECM) on aircraft platforms. Broadband, wide beamwidth, circularly polarized sensing capabilities were thus highly sought after. Interestingly, the discovery of an antenna capable of satisfying all those requirements came unexpectedly. In 1954, although discouraged by many experts, E. M. Turner wound a long-wire dipole into a spiral form and connected its terminals to a two-wire feedline.⁹ At that time, the largest antenna bandwidths were on the order of one octave, but the results obtained with the first spiral experiment were so encouraging that an immediate research effort was launched. At the present time, spirals and other frequency independent antennas are irreplaceable components of many communication platforms, various electronic warfare, ECM, and direction-finding systems, and atmosphere, ground, and space exploration stations, and so on.

Spiral antennas are inherently circularly polarized radiators with relatively constant input impedance and radiation patterns over wide frequency ranges. Bandwidth is determined by the fine precision of the feeding region (high frequencies) and overall spiral aperture size (low frequencies). Often the balun feed limits antenna performance, though very wideband feeds have been devised. Spirals can be designed in various planar or conical shapes, with the most common configurations being the equiangular^{2,10,11} and Archimedean spirals.¹² The finite features of spiral aperture prohibit its frequency independent operation for wavelengths greater than the antenna aperture circumference and less than the diameter of its feeding region. Spiral antennas operate in a so-called *3W* way—as a fast Wave, leaky Wave, and traveling Wave antenna. Basically, the *traveling* wave form of the excited currents on the spiral arms allows for its broadband performance. This wave is *fast* due to the mutual coupling between neighboring arms and *leaks* energy while propagating on the line to produce radiation.

Spirals can be designed with different numbers of arms, but typically two or four arms are used. The number of arms and their excitation phase/amplitude difference determines antenna performance, which is frequently described using spiral modes. Until recently, it was generally accepted that the number of broadband spiral modes that can be excited using an N -arm spiral is $N-1$ (see pages 108–109 of Corzine's and Mosko's *Four-Arm Spirals*¹³). However, in the article by Filipovic, Stutzke, Buck, Mu, and Kefauver, the broadband dual-mode operation of a two-arm slot spiral was successfully demonstrated.¹⁴ Two-arm spirals are most frequently used to operate in the 1st (mode 1, $m1$, sum, or Σ) mode, which is obtained when the arms are excited 180° out of phase. Radiation leakage occurs throughout the propagation path of the current traveling wave, with the most efficient radiation from the ring one wavelength in circumference. When operating in this mode, the spiral directs its radiation toward broadside.

Spiral antennas can be classified (with respect to the type of transmission line supporting the excited traveling wave) in two major groups: printed or wire spirals and slot spirals. Although the first published records of spiral research were on slot spirals,^{9–11} the majority of publications and patenting have occurred for printed geometries.^{4,5,12,13} The main reason is that multiarm designs are more easily realized with printed spirals. Additionally, once multi-octave broadband matching techniques were developed, vertical feeds were realized and many aperture designs became feasible. Other spiral designs have also been introduced, including the spiral mode microstrip (SMM) antennas, realized from microstrip lines wound into a spiral shape.¹⁵

Basic Principles

The *radiating ring theory*, also known as *band theory*,¹² is used to describe the theoretical principles behind the operation of spiral antennas. The band theory will be demonstrated on the simplest and most commonly used spiral antenna: a two-arm, planar, wire spiral antenna operating in mode 1. As depicted in Figure 13-6a, the spiral is fed from its center at ports A and A' , ideally with equal amplitudes and a 180° phase difference. When fed with an ideal balanced line, and assuming a discontinuity free aperture, the spiral will support propagation of the forward traveling wave. Mode 1 radiation will predominantly occur from a ring with approximately one guided wavelength (λ) circumference. When excited to operate in this mode, spiral currents at points B and B' , which belong to neighboring arms, will be directed the same way, i.e. they will have the same phase value. The same is true for diametrically opposite points C and C' . The nonradiated traveling wave currents will flow past this region, and if the size of the spiral permits, radiate in the next properly phased section. This will occur at a circumference equal to three guided wavelengths (mode 3 for a two-arm spiral). If the spiral is not large enough, the currents will reach the end of the spiral arms where they are either absorbed or reflected back toward the spiral's center. Note that if the size of the spiral is large enough, the in-phase current conditions will show up at odd guided wavelength circumferences of the spiral and higher order modes will radiate. This condition is known as *overmoding*. If attenuation of the forward traveling wave through radiation and dissipation in the lossy termination are insufficient, the currents will reflect back from the end of the spiral and backward traveling waves will be generated. The corresponding radiation of this current wave will be cross-polarized. The majority of backward current wave power will also leak in the odd wavelength circumference regions. Any remaining power will arrive at the spiral input ports (if perfectly matched) and show up as increased input VSWR. The phase of the traveling wave current is shown in Figure 13-6b, where two different shadings denote the instantaneous phases between $0-180^\circ$ and $180-0^\circ$. The mode regions with currents directed the same way are clearly seen.

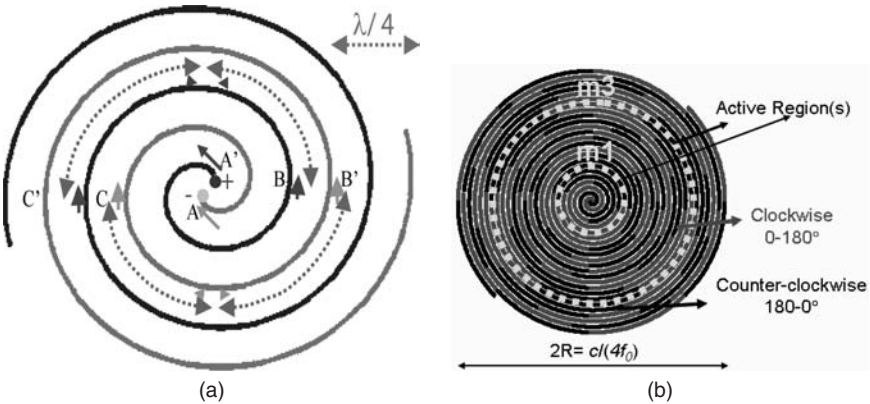


FIGURE 13-6 Band theory for mode 1 excitation (a) and current phasing (b)

A similar radiation process occurs with other modes and different numbers of arms. For example, a four-arm CCW spiral fed antenna in mode 1 (total phase progression of 360° at the input ports, i.e. the feed phase vector is $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$) will radiate higher modes of orders 5, 9, $4m + 1$, $m = 1, 2, \dots$ if the spiral is large enough to support those modes. When fed with 720° phase progression between input ports (feed-phase vector is $\{0^\circ, 180^\circ, 0^\circ, 180^\circ\}$), the spiral will operate in mode 2 (m_2 , difference or Δ), and overmoding will show up at radiating rings at 6, 10, $\dots 4m + 2$, $m = 1, 2, \dots$ guided wavelength circumferences. Mode 1 and 2 regions of a four-arm spiral are shown in Figure 13-7. Overmoding sometimes occurs due to imperfections in the feed circuitry, loose or improperly phased cables, material impurities, and other conditions capable of alternating the natural progression of the current traveling wave. This type of overmoding is not a function of the number of arms and is typically evenly distributed across the possible modes providing the errors are random. For the same four-arm spiral example used in Figure 13-7a, the errors of this type will be distributed across all modes (0, 2, 3, etc.).

Different Spiral Forms

The most frequently used spiral shapes are Archimedean and equiangular (log-periodic) spirals in planar or conical form. These shapes in planar form are drawn in Figure 13-8.

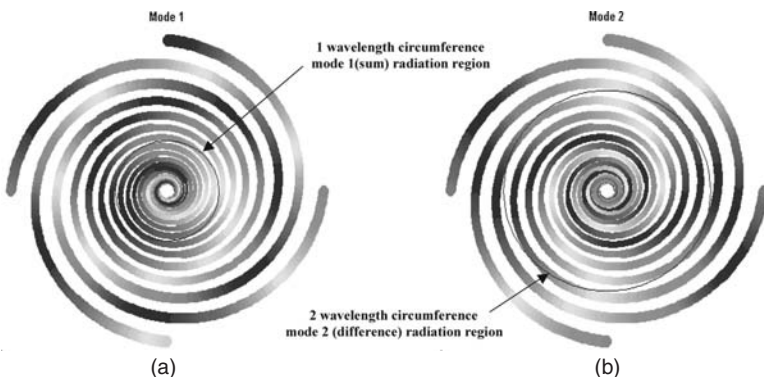


FIGURE 13-7 Four-arm spiral currents depicting mode 1 (a) and mode 2 (b) radiation regions

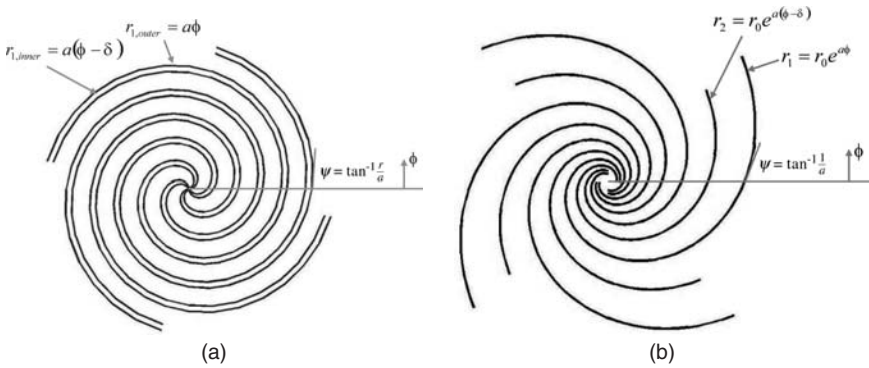


FIGURE 13-8 Four-arm Archimedean (a) and equiangular (b) spirals. Shown also are the equations used to generate the outline of their arms.

The Archimedean spiral has a constant arm width and constant separation between arms through the entire aperture. The defining equation for this antenna is $r = r_0 + a\phi$, where r_0 is the starting radius, a is the growth rate, and ϕ is the progressive growth angle. The equiangular spiral has progressively increased arm width and separation between the arms as they open toward the outside. This shape can be entirely described by angles and is the basis for the frequency independent principles (see Section 13.1). The defining equation for an equiangular spiral is $r = r_0 e^{a\phi}$, where r_0 is the starting radius, a is the growth rate, and ϕ is the progressive growth angle. Another term frequently used to describe tightness of the equiangular spirals is the so-called expansion factor (EXP), defined as $EXP = e^{2\pi a}$. For an expansion factor of 2, the spiral radius, as well as the arm width and separation between the arms, will double for each turn. Shown in Figure 13-8 are also the equations for the angles subtended between the radius vector and the tangent on the spiral line at the same point along the arms. For the Archimedean spiral, this angle will depend on the radius, r , while with the equiangular structure, this angle is constant.

Two-arm Spirals

The minimum number of arms needed for single-mode broadband operation is two. Single arm spirals have been designed for some narrow-band applications and will not be discussed in this chapter.¹⁶ The most frequently used two-arm spirals are two-arm Archimedean spirals. A commercial version of this antenna is shown in Figure 13-9. These 5 cm diameter spirals work from 2 to 18 GHz and are available in both RCP and LCP versions. The depth of the cavity backing is almost as large as the antenna diameter. About half of this depth is due to absorber loading and the remaining is due to the integrated balun. The absorber loading significantly reduces the interaction between the metal backing and the spiral aperture. Use of absorber loading reduces antenna efficiency to a maximum of about 50 percent; however, the pattern uniformity over the operating bandwidth is that of a freestanding structure. Integrated within the cavity is also a vertical balun/impedance transformer connected between the spiral ports (in the center) and the connector on the back. The copolarized and cross-polarized gains of this antenna vary little, and the axial ratio is better than 1 dB throughout the band. The axial ratio is highly dependent on the termination method used, especially toward the low end of the operating band (near one wavelength circumference for a mode 1 spiral).

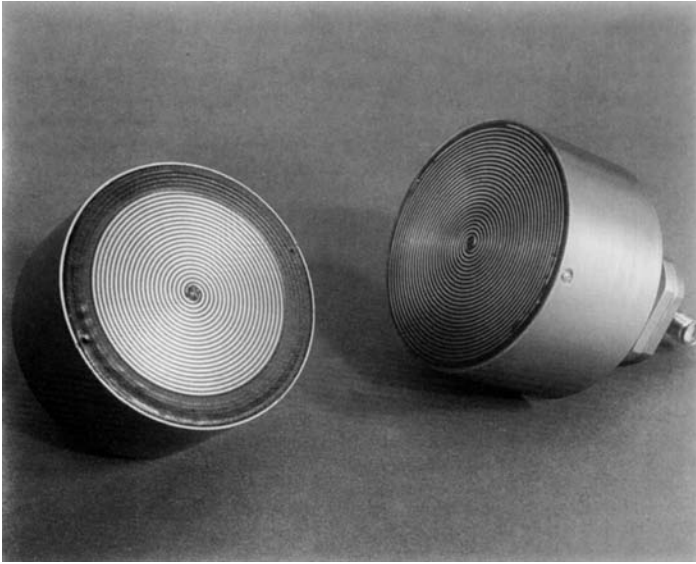


FIGURE 13-9 Two 5-cm diameter complementary Archimedean spirals terminated with absorptive paint and a 100- Ω resistor; arm width $W \approx 0.58$ mm (Courtesy Loral Randtron Systems)

The typical impedance and far-field performances of a well-designed 5-cm diameter two-arm freestanding spiral without end termination are shown in Figure 13-10. The low-frequency ripples in impedance and high axial ratio denote the effects of the backward wave and are the consequence of unterminated arm ends. At the low-frequency end, the diameter of the antenna is not large enough to support the radiation of the $m1$ through the entire active region, thus a strong backward wave exists on the arms. When this wave

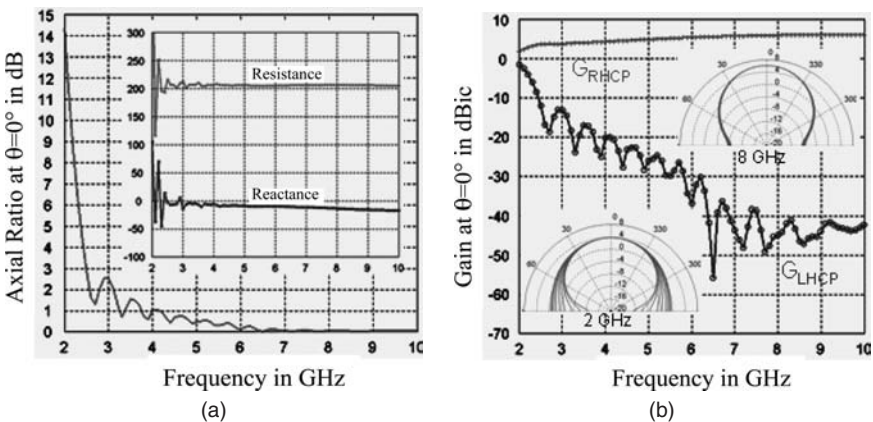


FIGURE 13-10 Broadside axial ratio and input impedance (a), and co/cross-polarized broadside gains with overlaid 36 azimuthal cuts at 2 and 8 GHz (b) of a free-standing self-complementary two-arm spiral. These results are obtained with the method of moments code FEKO.

couples to the feed, it deteriorates the antenna impedance. The radiated power of the backward wave is high, so that the two circularly polarized patterns are superimposed, giving rise to the linearly polarized performance. Additionally, the omnidirectionality deteriorates at 2 GHz as seen from the spread of 36 overlaid copolar patterns shown in the inset of Figure 13-10b. At 8 GHz, this spread is significantly smaller, and a symmetric radiation pattern is obtained.

Several variations of two-arm spirals are shown in Figure 13-11. Different metal to nonmetal (slot) ratios are typically used for tuning the impedance at the antenna ports. However, the most uniform far-field performance is obtained with self-complementary spirals. The differences in performance between the tightly wound Archimedean and equiangular spiral are not significant, and sometimes a combined growth is utilized. Actually, the far-field of tightly wound Archimedean and equiangular planar spirals is undistinguishable. However, for conical projections, the differences in beamwidths can be measured.¹⁷

Multiarm Spirals

The interest in spirals with more than two arms was ignited by Paul Shelton in 1960. He proposed three- and four-arm spiral designs for multimode operation (see Appendix A in *Four-Arm Spirals* for the copy of the original letter¹³). Shelton recognized that for proper phase progression at the input ports an extra arm (reference) was needed; thus to achieve two broadband modes, at least three arms were required. His finding discontinued the ongoing research regarding demonstration of two broadband modes from two-arm spirals.¹⁸ Since then, it has been generally accepted that an N -arm spiral can radiate only $N - 1$ broadband modes. Therefore, the minimum number of arms for a multimode spiral is three. A unique feature of the three-arm spiral is that the first and second modes have the same impedance. However, four arms are generally used due to the more complex feed network, often called *beamformer* or *modeformer*, needed for the three-arm spiral (feed phase vectors are $m_1 \{0, 120, 240^\circ\}$ and $m_2 \{0, 240, 480^\circ\}$). For dual-polarized spirals, such as the *Modulated Arm Width (MAW)* spiral,¹⁹ a minimum of five arms are required for multimode operation. Due to beamformer simplicity, six or eight arms are generally used. Shown in Figure 13-12 are wire models of two-, four- and six- arm log-spirals with the same exponential factor $EXP = 1.5$. Starting radius is 1.25 mm; ending radius is 3 cm.

When comparing the far-field performance of these antennas, the major differences are in the symmetry of the azimuthal patterns (WoW, which will be discussed more in the next section) and in the broadside axial ratio (cross-polarization). The copolarized patterns are virtually indistinguishable. The complexity of the mode-forming network and arm terminations are the main reasons for the lesser use of spirals with more than two arms.

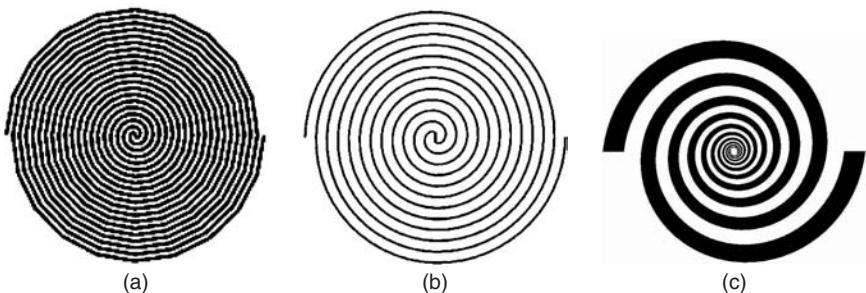


FIGURE 13-11 Drawing of a self-complementary Archimedean (a), non-self-complementary Archimedean spiral (b), and self-complementary equiangular (c) two-arm spirals

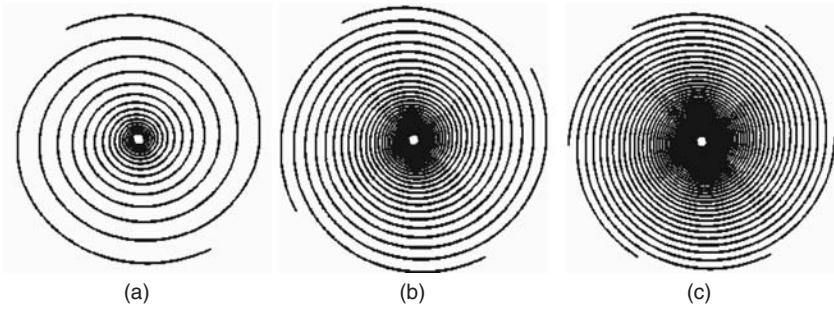


FIGURE 13-12 Two- (a), four- (b), and six-arm (c) equiangular spirals with $EXP = 1.5$

Shown in Figure 13-13 are the cross-polarization levels and the 3-dB beamwidth for two-, four-, and six-arm spiral antennas when designed for mode 1 operation. Spirals are free-standing, and unterminated arms are made of a perfect electric conductor. The increased number of arms contributes to the improved cross-polarization levels, primarily due to the increased coupling in the active region, and better filtering of negative (cross-pol) modes. The copolarized patterns are unaffected by the number of arms used.

Spiral Modes

Spiral modes were first theoretically studied in the early 1960s^{20,21} when radiation conditions for orthogonal radiation modes of infinite arm spirals were derived. It was shown that an arbitrary feeding arrangement at the spiral arms can be decomposed into the sum of several modes, giving rise to the characteristic far-field radiation. Multimode spirals have been traditionally used for monopulse direction finding,¹³ where amplitude and phase comparisons between the multiple modes can provide information on azimuth and elevation of the incoming signal.

The notion of spiral modes is associated with the desired or undesired radiation conditions from a spiral aperture. Desired radiation is that of the dominant modes for which the spiral is designed. Undesired radiation is that occurring from various imperfections (beamformer errors, phase matching from the cables or balun, and so on) or excitation

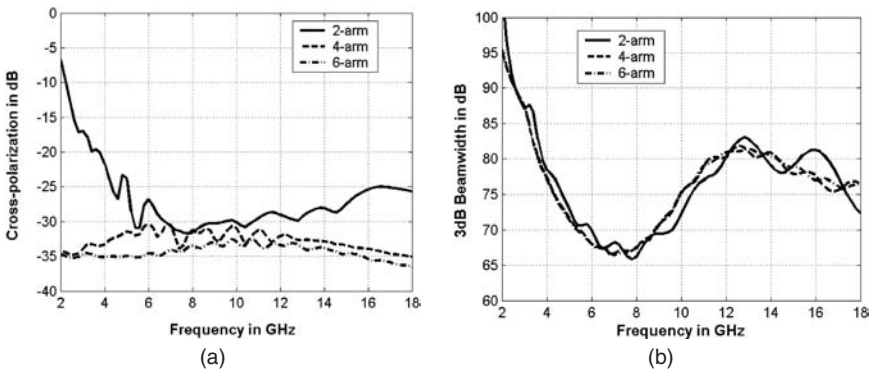


FIGURE 13-13 Cross-polarization difference (a) and 3 dB beamwidth (b) of two-, four-, and six-arm spirals

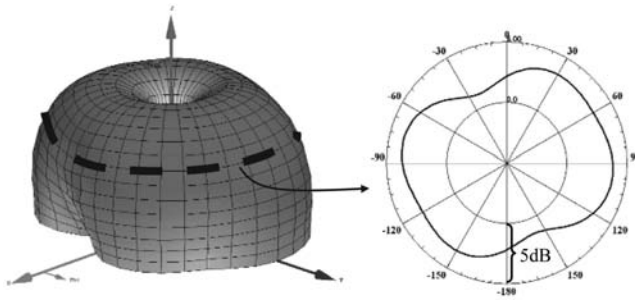


FIGURE 13-14 Effect of feed imperfections on the mode 2 radiation pattern of a four-arm spiral

and radiation from the higher order modes. For example, for a four-arm spiral designed to operate in mode 2, the imperfections can cause radiation from modes 1, 3, 4, etc., while the higher-order mode is mode 6. An example of an undesired mode 4 effect on the mode 2 pattern of a four-arm spiral is shown in Figure 13-14. The pattern undulation with two half sinusoids occurred due to the imperfections in the feeds.

Dominant modes are those for which the proper phase progressions at the feed are devised; for mode m , an overall phase progression of $m \times 360^\circ$ is needed. They are circularly polarized waves radiated from the $m\lambda$ circumference rings and have a far-field phase progression with m cycles of 360° on any constant θ angle. Only mode 1, both right circular polarization (RCP) and left circular polarization (LCP), has any boresight component. All other modes have an on-axis null. Higher-order modes are typically unwanted, and they are formed due to the existence of the nonradiated currents on the spiral aperture outside the nominal radiating ring. They can be suppressed by reducing the size of the spiral so its circumference is smaller than $m\lambda$, where m is the unwanted mode. Improving the efficiency of the radiation through the dominant radiating ring can reduce the negative effects of these modes, and this is recommended for spiral designers. Radiated power of the higher-order modes is to blame for scalloping spiral patterns and significantly increased amplitude wobble on conical spiral pattern cuts (the parameter often termed as $WoW^{5,13}$). Shown in Figure 13-15 are the decomposed patterns of desired mode 1 and higher-order modes 3 and 5 of a two-arm spiral on a constant conical cut. The undulations associated with the higher-order modes clearly verify the previous discussion, and special care must be exercised when truly frequency independent performance is needed of a spiral.

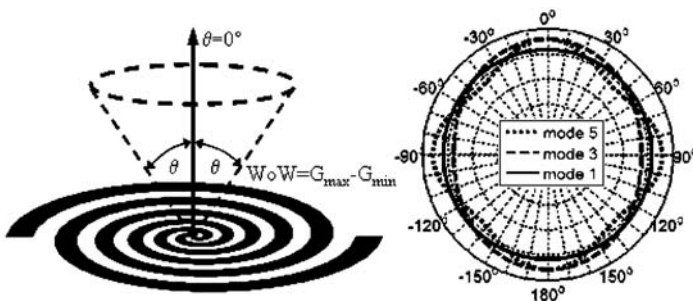


FIGURE 13-15 Azimuthal patterns for mode 1 and higher-order modes 3 and 5 superimposed with mode 1

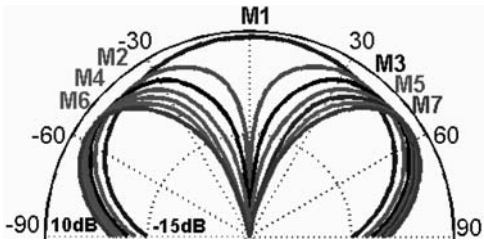


FIGURE 13-16 Radiation patterns of modes *m*1 through *m*7 for an eight-arm freestanding spiral

Computed copolarized radiation patterns for seven broadband modes of an eight-arm equiangular spiral are shown in Figure 13-16. Each higher-order spiral mode peaks successively farther from boresight with only mode 1 having maximum gain at boresight. Approximate beam peak angles and directivities are shown in Table 13-1. These values are typical but there will be minor variations based on spiral design and frequency.

To demonstrate overmoding on a two-arm spiral antenna, the max gains of the operating mode *m*1 and a first higher-order mode, *m*3, are computed for several equiangular spirals with various growth rates and their difference is plotted in Figure 13-17. The tighter growth rates can be utilized to design antennas with high cross-mode difference gains and thus improved patterns. The maximum value shown for *m*3 radiation is –18 dB below the *m*1 gain at 18 GHz.

The WoW at 30° for the antennas given in Figure 13-12 is shown in Figure 13-18. Spirals with a larger number of arms produce more symmetric (pure) radiation patterns due to the progressively higher order of the first contaminating mode. As discussed earlier, for the two-arm spiral, the first higher-order mode is *m*3, while for the four- and six-arm spirals, these are modes *m*5 and *m*7, respectively. Unfortunately, the complexity of the beamformer and the arm terminations prevent the widespread use of spirals with more than two arms.

Dielectric Loading Effects

A spiral is a fast wave antenna and the use of a high dielectric constant and/or thick dielectric can significantly alter its characteristics. Input resistance is reduced and the gain, axial ratio, and pattern purity in general are all degraded when compared with a freestanding spiral. Dielectric loading slows the traveling wave thus reducing the aperture of the active mode. Additionally, the coupling between the neighboring arms is reduced and the radiation through the active region is decreased. This means that the forward traveling wave, after passing the desired region, will have more energy and radiate in the higher order modes, contributing to the excessive far-field contamination. For some applications, pattern purity is

TABLE 13-1 Directivity and Beam Peak Elevation Angles for Seven Modes of an Eight-arm Freestanding Spiral

Spiral mode	<i>m</i> 1	<i>m</i> 2	<i>m</i> 3	<i>m</i> 4	<i>m</i> 5	<i>m</i> 6	<i>m</i> 7
Beam peak in°	0	36	48	52	54	56	57
Max directivity in dBic	8.5	6.3	6.3	6.5	6.7	6.8	6.8

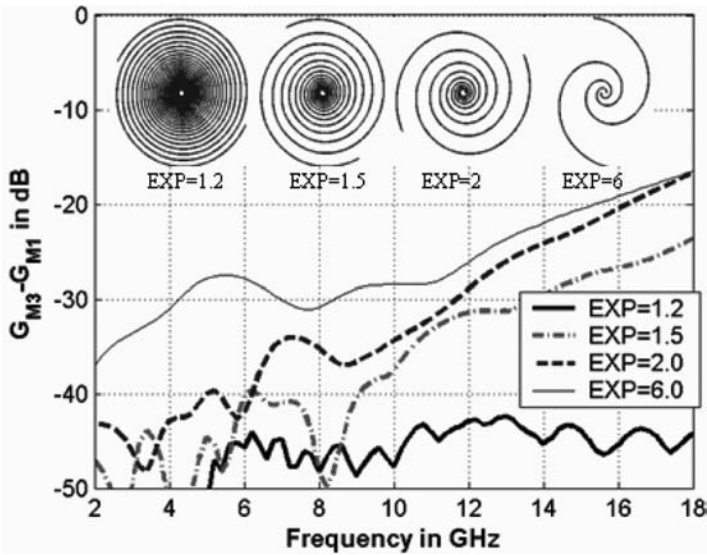


FIGURE 13-17 Mode 3 content (relative to mode 1) of a 6-cm diameter two-arm equiangular spiral with various growth rates shown in inset

not essential so that the associated miniaturization due to the loading is usually computed as a relative reduction of the lowest frequency for which the spiral has a prescribed gain value at the beam peak.²² Pure modal patterns can still be obtained, but the number of spiral arms need to be increased or the operational bandwidth is decreased (Eq. 13-15). Interestingly, and owing to the radiation ring theory of operation, this translation of the frequency dependent gain curve is proportional to $(\epsilon_{r,\text{eff}})^{1/4}$, as compared to the $(\epsilon_{r,\text{eff}})^{1/2}$ impedance shift, where $\epsilon_{r,\text{eff}}$ is the effective dielectric constant of the associated stripline model.²³

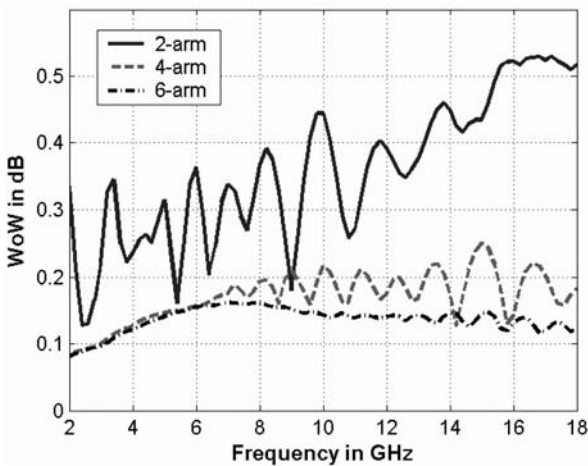


FIGURE 13-18 WoW at 30° for spirals given in Figure 13-12

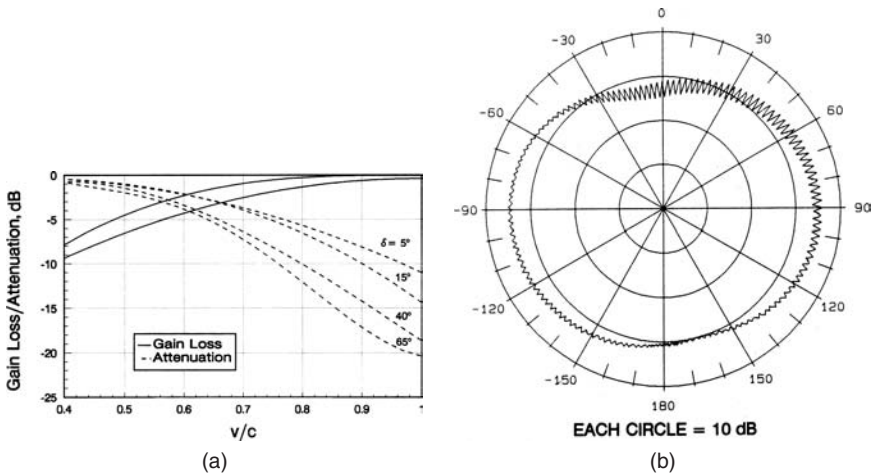


FIGURE 13-19 (a) One λ -ring gain loss and attenuation versus v/c for an equiangular spiral. (b) Measured rotating linear conical-cut pattern of a 5 cm cavity-backed Archimedean spiral with $v/c = 0.65$ (Courtesy of Loral Randtron Systems)

Effects of dielectric loading are best shown in Figure 13-19. The free-space case has the lowest gain reduction and highest attenuation of the traveling wave through the radiating region. The more pronounced the loading (lower wave velocity v), the more bound the traveling wave is to the spiral and the less it radiates. The conical-cut pattern of a spiral with $v/c = 0.65$ taken at the $\theta = 60^\circ$ clearly depicts the undesired loading effects on the pattern purity. The WoW of about 8 dB, the variation of the axial ratio from 0.1 to 4 dB, and two undulations offer clear proof of a very high-level of energy being radiated from the $m3$ ring, i.e. reduced attenuation through the $m1$ active ring.

Far-Field Modal Decomposition

To investigate the composition of the modes in a radiation pattern of a spiral antenna, a technique referred to as *pattern modal decomposition* is often used.^{5,13,24} The radiating modes of an N -arm spiral can be obtained by producing a phase progression between adjacent arms at the feed points equal to $2\pi m/N$, where $m = 1, 2, \dots$ is the desired mode of operation. To compute the modal content of the radiated field, the phase progression is applied to the arms of the spiral through complex voltage coefficients defined by

$$V_{k,m} = V_{0m} \cdot e^{\frac{-j2\pi m(k-1)}{N}} \quad (13-8)$$

where $k = 1, \dots, N$ denotes the arm, V_{0m} is a constant amplitude, and j equals $\sqrt{-1}$. Equation 13-9 has the restriction

$$\sum_{k=1}^N V_{k,m} = 0 \quad (13-9)$$

for each mode. A determination of the modal fields in terms of the measured pattern is then obtained by applying

$$E_m(\theta) \cong \frac{1}{N_\phi} \cdot \sum_{k=1}^{N_\phi} V_{k,m}^* \cdot E(\theta, \phi_k) \quad (13-10)$$

where $E_m(\theta)$ are the coefficients to the discrete Fourier series representation of $E(\theta, \phi_k)$. N_ϕ is the number of points per conic cut (at a constant theta value), in the far-field pattern data, and the highest number of ambiguity free modes that can be resolved is equal to $N_\phi/2-1$. $E(\theta, \phi_k)$ can be measured pattern data or calculated from electromagnetic models.

To demonstrate the capabilities of this technique, the beamformed patterns of an eight-arm RCP spiral are decomposed into modes, as shown in Figure 13-20. Both left- and right-handed modes supported by this structure are measured, and the power in cross-polarized modes (negative ones) is much less than that of the copolarized fields. Also clear from this figure is that the higher-order modes “turn on” at progressively higher frequencies. A single arm measurement without beamformer can also be decomposed into basic spiral modes, thus providing a significant amount of data, i.e., all possible spiral modes from a single port measurement.

Mode Impedances

The input impedance to the ground for a self-complementary N -arm structure in free space is derived by Deschamps²⁵ and is given as

$$Z_m^{\text{fs,comp}} = \frac{\eta_0/4}{\sin\left(\frac{m \cdot \pi}{N}\right)} \quad (13-11)$$

where $\eta_0 = 120\pi \Omega$ is the impedance of the free space. The integer, m , denotes the eigenvalue or characteristic mode ($m = 1, 2, \dots, N-1$) where N is the number of arms on the structure.

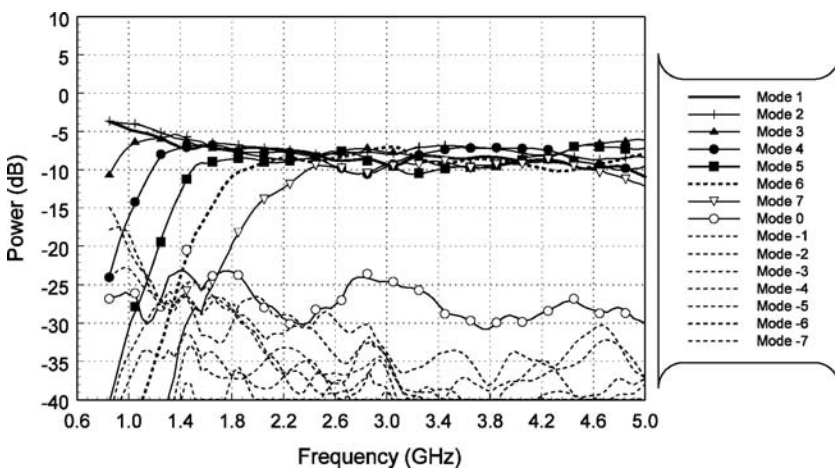


FIGURE 13-20 Power in each mode for a single arm of a 25-cm diameter eight-arm spiral

TABLE 13-2 Input Impedance of Free-space Self-Complementary Spirals

# Arms	Spiral Mode						
	1	2	3	4	5	6	7
2	94.2						
3	108.8	108.8					
4	133.3	94.2	133.3				
5	160.3	99.1	99.1	160.3			
6	188.5	108.8	94.2	108.8	188.5		
7	217.2	120.5	96.7	96.7	120.5	217.2	
8	246.3	133.3	102.0	94.2	102.0	133.3	246.3

Modal impedances for self-complementary two- to eight-arm spirals are computed using Eq. 13-11 and the results are shown in Table 13-2. Interestingly, only a three-arm spiral has the same modal impedances for $m1$ and $m2$, making it an excellent candidate for monopulse direction finding. However, the complexity of the beamforming network for this antenna reduces its practical application. In practice, these theoretical impedances are lowered due to the finite thickness of the metallic parts on an antenna as well as the inevitable dielectric loading (used for mechanical support).

To impedance match the spiral to a particular mode, the arm to gap ratio (A/G) at the spiral feed points can be adjusted to match to the feed cable impedance (usually 50Ω). Because spirals radiate most efficiently as self-complementary structures, the noncomplementary feed region should be slowly transformed back to the complementary before the first desired radiation region. As with most impedance transformers, this transition should be at least one-quarter guided wavelength long at the lowest frequency of operation. A drawing of a tapered A/G feed region for an eight-arm spiral is shown in Figure 13-21*a*. For this antenna to simultaneously match both m_2 and m_3 to 50Ω coax, the predicted optimum A/G ratio is about 20 (see Figure 13-22). Impedances for various A/G ratios, modes, and dielectric thickness are computed by Huffman and Cencich.²⁶

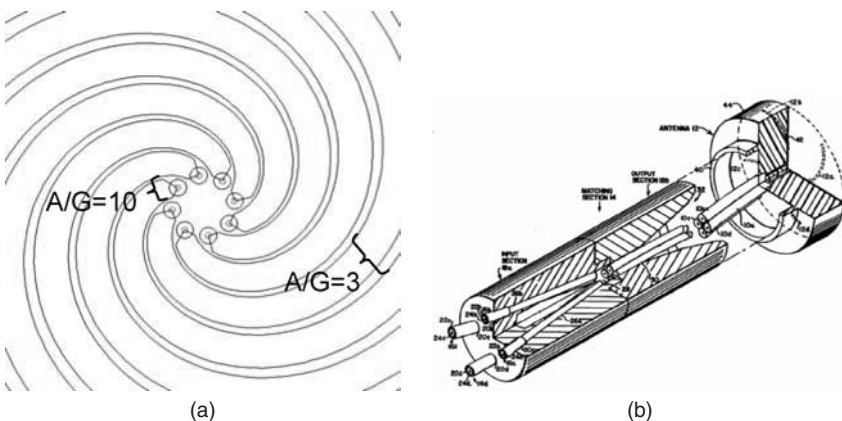


FIGURE 13-21 (a) Impedance transformer for an eight-arm spiral and (b) matching section for a four-arm spiral antenna²⁸

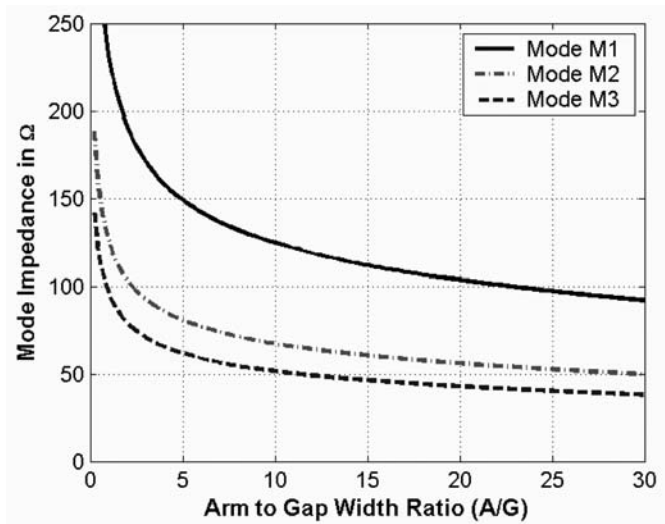


FIGURE 13-22 Theoretical input impedance versus A/G width ratio for modes 1, 2, and 3 of an eight-arm spiral antenna on a 0.01-inch-thick 2.1 dielectric constant substrate

Simultaneous matching of spiral mode impedances is a challenging problem and several methods have been devised. These methods utilize a feedline structure that is highly decoupled at the beamformer input (all input ports are 50Ω , regardless of mode), and strongly coupled at the spiral interface (impedance is dependent on input excitation). In the feedline section, near the spiral, coupling is obtained similar to that between the arms in a spiral antenna. A six-arm spiral with a coaxial-type coupled feedline²⁷ has maximum return loss of -13 , -17 and -16 dB for the first three modes. Average return loss over the 4:1 bandwidth was -18 , -22 and -24 dB for these same modes. Another approach to simultaneously match both sum and difference modes of a four-arm spiral is shown in Figure 13-21*b*.

Spiral Feeding Techniques

A number of different feeding techniques have been developed for spiral antennas. Some include not only balanced to unbalanced transitions but also impedance transformation. The tapered microstrip balun,²⁹ Marchand balun,³⁰ and 180-degree hybrids with coaxial feeds are commonly used for two-arm printed spirals. The tapered microstrip balun, shown in Figure 13-23*a*, has the advantage that the spiral impedance can be matched by tapering the width of the strips. It can be designed for bandwidths several octaves wide. An infinite (Dyson) balun feed¹⁰ has theoretically unlimited bandwidth and is the most frequently used horizontal feeding technique. The outer conductor of the coaxial cables is soldered to the metallic portion of the spiral (in between the slot arms), and a transition to the slotline is provided in the center. A two-arm slot spiral with this balun is shown in Figure 13-23*b*. The dummy cable is used to provide the geometrical symmetry; however, if not used, the beam squint is not that significant. As shown in Figure 13-23*c*, a vertical single coaxial feed can be easily implemented with a four-arm spiral operating in mode 2.³¹ Finally, the most general feeding arrangement for spirals with more than two arms is shown in Figure 13-23*d*. Here, coaxial feedlines are used, where the outer conductors of these coaxial cables are soldered together to form a compact coaxial cluster. To further reduce unwanted feed line radiation,

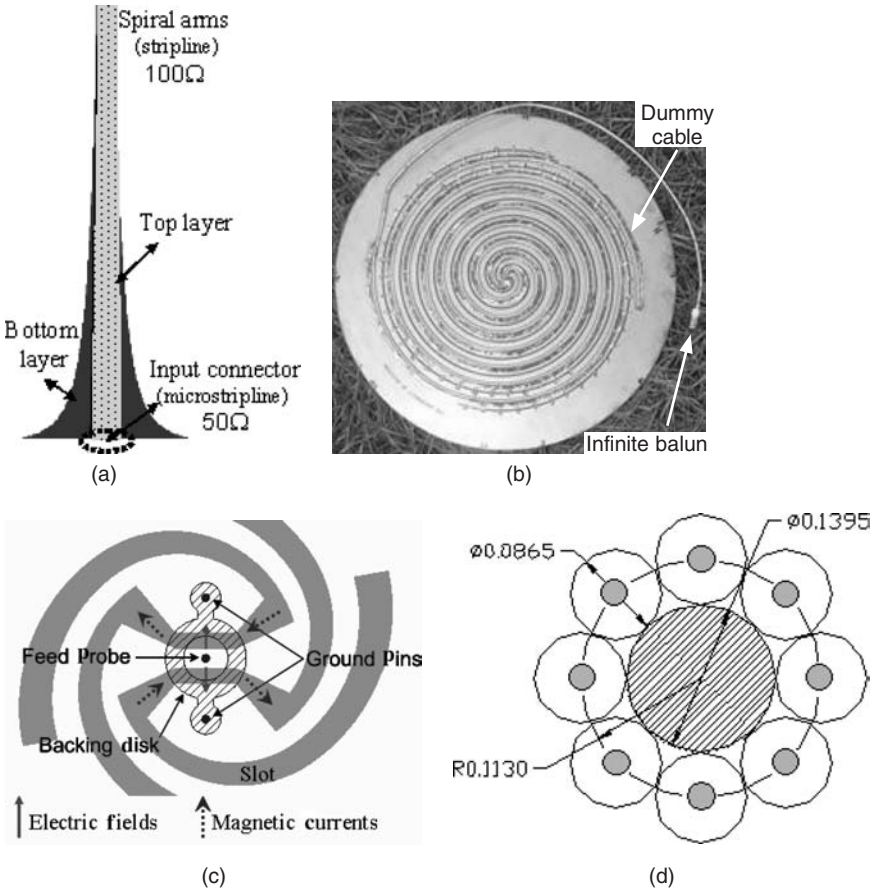


FIGURE 13-23 (a) Parallel strips balun/transformer, (b) infinite coaxial Dyson balun, (c) vertical coaxial balun, and (d) coaxial cluster feed

absorber paint is sometimes used to coat the outer conductor of the coaxial feedline cables between the beamformer and spiral.

If the spiral start radius is too large, excessive energy will leak through the mode 1 radiation region. Spiral size permitting, this energy will radiate in higher-order modes, thus significantly contaminating its far-field. The starting radius guideline for spirals with relatively small growth rates ($EXP < 2$) is

$$R_{\text{start}} = \frac{\lambda_{\text{high}} m_{\text{lowest}}}{6\pi} \quad (13-12)$$

where λ_{high} is the guided wavelength of the highest frequency of operation and m_{lowest} is the mode order of the lowest desired mode of operation. Additionally, the feed region of the spiral can interact with the feeding cables and create currents that propagate on the outside of the cables. This can severely deteriorate the radiation patterns and also provide unwanted interference. Since multiarm spirals are commonly fed with some form of coaxial clusters, special care must be exercised when the minimum spiral radius is determined.

TABLE 13-3 Maximum Mode 1 Frequency for Coaxial Bundle Feed (for commercially available cables)

Arms/Coax	Maximum frequency of operation in GHz			
	0.141"	0.085"	0.047"	0.034"
2	7	12	22	30
4	6	10	18	25
6	5	8	14	20
8	4	6	12	16

The approximate upper frequency limits for quality mode 1 radiated patterns for various coaxial cable sizes are summarized in Table 13-3.

As discussed earlier, a spiral needs to be $m\lambda$ in circumference to support the radiation of mode m . Since the radiation region is actually a finite width ring, it will be severely truncated if the circumference is only $m\lambda$. A good rule to avoid unnecessary truncation of the radiating ring of a tightly wound spiral ($\text{EXP} < 2$) is expressed in

$$D \geq \frac{(m_{\text{highest}} + 0.5)\lambda_{\text{low}}}{\pi} \tag{13-13}$$

where D is the spiral diameter, λ_{low} is the guided wavelength of the lowest frequency of operation, and m_{highest} is the mode order of the highest desired mode of operation. For higher spiral efficiency, the numerator of Eq. 13-13 should be increased to $(m_{\text{highest}} + 1)\lambda_{\text{low}}$. Loss (directivity minus gain) versus spiral circumference for a four-arm printed equiangular spiral over an absorptive cavity is shown in Figure 13-24. At least 3 dB of this loss can be attributed to absorption of the back hemisphere pattern of the spiral.

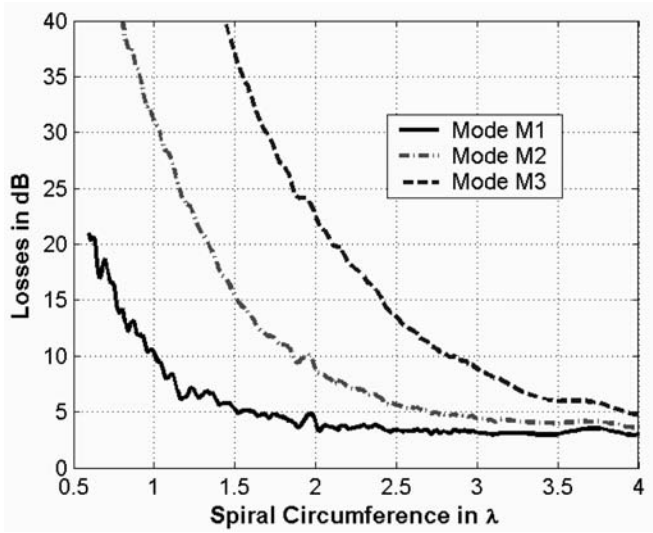


FIGURE 13-24 Difference between directivity and gain for broadband modes of a four-arm printed spiral backed by an absorptive cavity

Spiral Backing Techniques

A planar spiral antenna radiates bidirectionally, with opposite polarization senses in the two hemispheres. For most applications unidirectional radiation is required. The most commonly utilized approaches for eliminating the backward radiation of a planar spiral are use of either a reflective- or absorber-filled cavity. Absorptive cavities dissipate backward radiation using carbon-filled materials in some form of honeycomb, syntactic foam, or polyurethane foam. Typical configuration of an absorber-backed spiral is shown in Figure 13-25. A graded absorber is used to taper match the air to the heavily loaded carbon material. If the absorber is placed too close to the spiral element, additional attenuation of the forward traveling wave will occur contributing to increased losses (greater than 3 dB). A dielectric lens placed on the top of the spiral can be used to recover some losses (~ 1 dB) due to the redistribution of the radiated power in two hemispheres.

Computed effects of the perfect electric conductor (PEC) or perfect magnetic conductor (PMC) backings on the WoW of a good two-arm spiral antenna are shown in Figure 13-26.³² Neither of the two backing configurations can provide uncontaminated far-fields over a broad frequency range, although they can be utilized over limited bandwidths. Compared to the free-space (or effectively well-designed absorber-backed antenna), the pattern asymmetries are excessive. In the case of PEC backing, the fields reflected off the metallic bottom undergo polarization reversal and constructively add in the far-field (at broadside) if the bottom

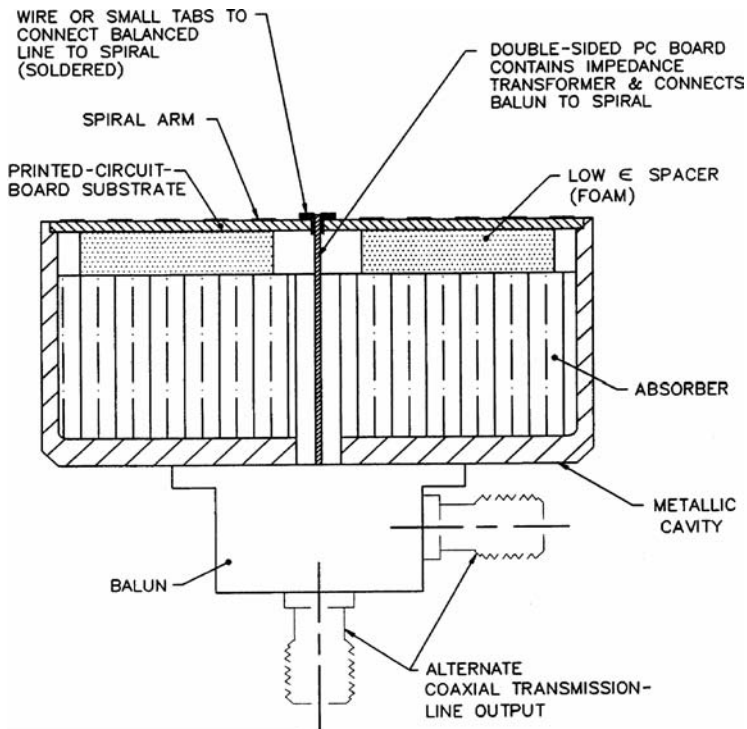


FIGURE 13-25 Sketch of an absorber backed cavity along with accompanying parts of a practical spiral

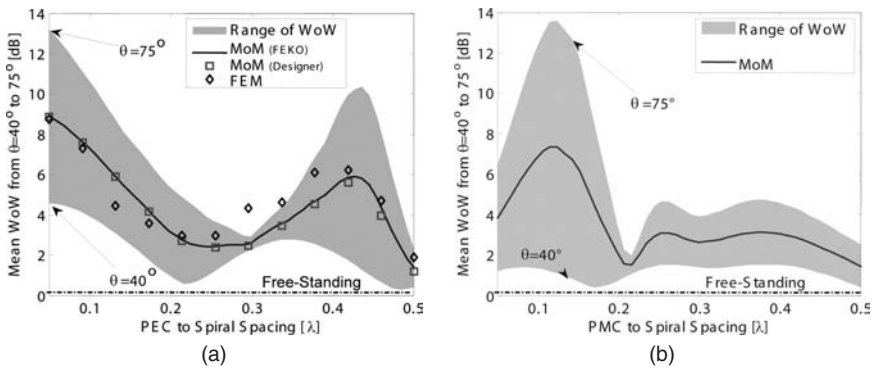


FIGURE 13-26 The effect of infinite PEC (a) and PMC (b) backings (at different depths) can be observed on WoW and azimuthal pattern symmetry. The mean value of WoW from $\theta = 40^\circ$ to 75° is plotted with the range over the same shaded elevation span. The absorber-backed or freestanding value is also included. The WoW of the freestanding structure varies from 0.08 dB at $\theta = 40^\circ$ to 0.3 dB at $\theta = 75^\circ$ (after M. Buck and D. Filipovic³² © IEEE 2006).

is a quarter wavelength from the spiral. Overmoding will occur if the spiral size permits creation of higher-order mode active regions. Thus, an effective way for suppressing overmoding effects is to increase the number of spiral arms such that the active region of the lowest-order parasitic mode falls outside of the spiral aperture. For example, if you want to design a mode 1 spiral to operate in 2–20GHz range with a 5 cm aperture diameter and absorber-free metallic backing, how many arms should you consider? Assuming ideal conditions on the spiral (zero width of the active region, no miniaturization, and ideal arms termination) and knowing that the circumference of the active region for the first parasitic mode is $(N + 1)\lambda_{\text{high}}$, you can compute the number of arms as $N \geq \pi/\lambda_{\text{high}} - 1$, i.e., at least $N = 10$ arms are needed. A more appropriate inequality for computing the number of arms for this type of spiral antenna is

$$N = \left\lfloor \frac{f_{\text{high}}}{f_{\text{low}}} (m + \Delta) \right\rfloor - m + 1 \quad (13-14)$$

where f_{high} and f_{low} are highest and lowest frequency of operation, m is the desired mode of operation ($m = 1$ in the given example), $\lfloor x \rfloor$ is the integer part of x , and Δ is the safety factor depending on the width of the active region, depth of metallic backing, and established criteria for tolerable cross-modal difference (typical values for Δ are ~0.5 to 1). Note that multiarm designs have higher complexity and price associated primarily with the beamformer. Often the increase in antenna efficiency does not compensate for the excessive cost (remember that the beamformer must have the same bandwidth as the antenna; thus for a 12-arm spiral maintaining 30° phase progression between neighboring ports, over a decade of bandwidth is a very difficult task).

Effects of the metal backing are also shown in Figure 13-27, where modal content of metal backed and free-standing two-arm spirals are compared. The power radiated by the higher-order modes of the free-standing spiral is significantly lower. Note that the cross-modal difference of 10 dB typically causes WoW on the order of 6 dB on wide-angle conical cuts. A metallic backing spaced farther than $\lambda/4$ will begin to degrade the boresight gain. At a height of $\lambda/2$, the mode 1 pattern will have a null on boresight but still maintain mode 1 phase characteristics.

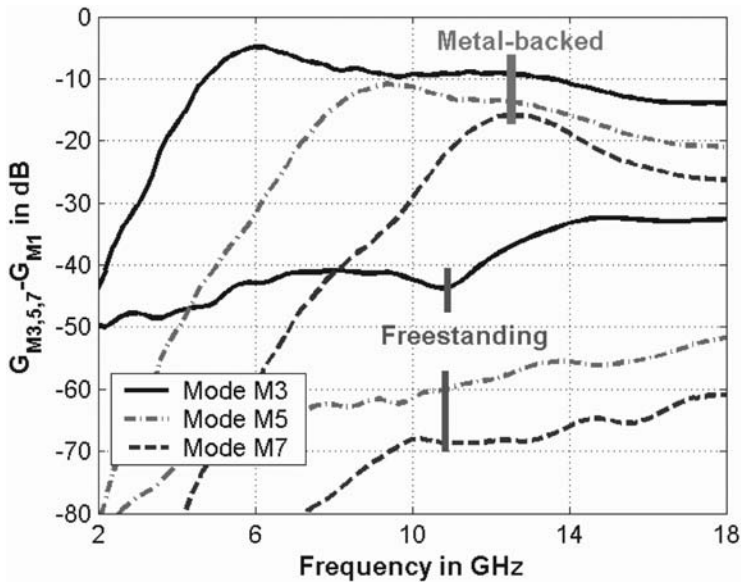


FIGURE 13-27 Modal contamination of cavity-backed and free-standing two-arm spiral

Arm Termination Techniques

When attenuation through the active region(s) is insufficient, the residual traveling wave will reach the arm ends, encounter a discontinuity (open end for printed and short end for slot spirals), and reflect back toward the feed. Simultaneous existence of two oppositely directed traveling waves on the spiral create conditions for a standing wave, thus increased VSWR and crosspol. To mitigate these effects, special care must be taken in the design of arm terminations. Various techniques for terminating the spiral arms have been devised in the past including

- Distributed resistive loading in the form of matched tapers such as exponential or Klopfenstein tapers (preferred method for slot spirals) (see Figure 13-23*b*, and for a detailed design, see Nurnberger's and Volakis's paper³³)
- Absorber paint on the last turn (or on a few turns, see Figure 13-28*a* with ambidextrous spiral configuration)
- Resistive termination to the spiral cavity, as shown in Figure 13-28*b*
- Single resistor loaded microstrip or stripline transition or a high-impedance slot circle
- Absorber ring around the spiral perimeter, either on the top (contact) or inside the cavity
- Helix-loaded spiral (absorber loads the helix at low frequencies, see Figure 13-32)
- A thin (thus lossy) multiple turn wire at the spiral perimeter, sometimes painted with lossy material
- Multiple turns of a long, thin arm's extension (see Figure 13-28*c*)

The rule of thumb is that the termination section composed of distributed elements should be at least a quarter guided wavelength long at the lowest frequency of the spiral.

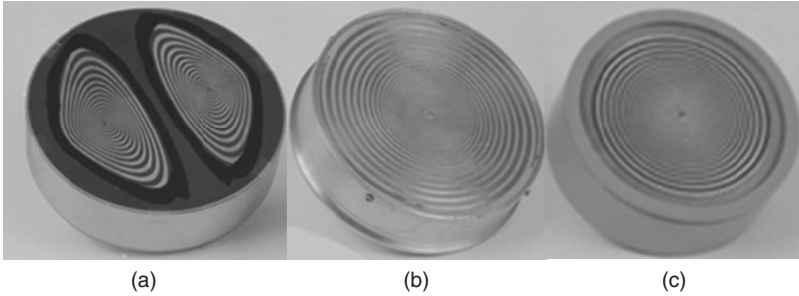


FIGURE 13-28 Photos of several spiral terminations: (a) absorber paint, (b) resistor, and (c) thin arm (Courtesy of BAE Systems)

Conical Spirals

To reduce the backward radiation of a free-standing planar spiral, Dyson wrapped the spiral onto a cone.¹¹ Parameters needed for constructing a conical spiral and a picture of this type of antenna is shown in Figure 13-29. This form can increase radiation toward the cone apex and reduce back hemisphere radiation. Printed and slot designs, including spirals with outer coaxial conductors playing the role of arms, have been realized, and they were fed with some form of Dyson balun. Dyson determined that when the spiral currents are reduced by 3 dB on the high-frequency end (at the top of the cone) and 15 dB (at the bottom of the cone) that the

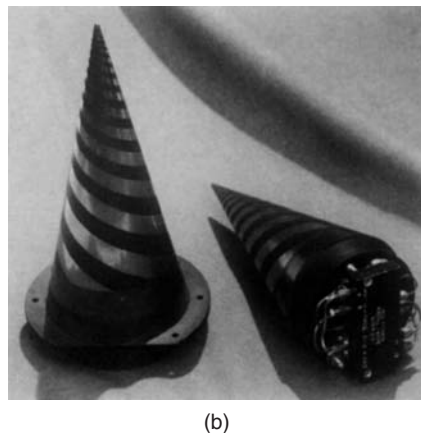
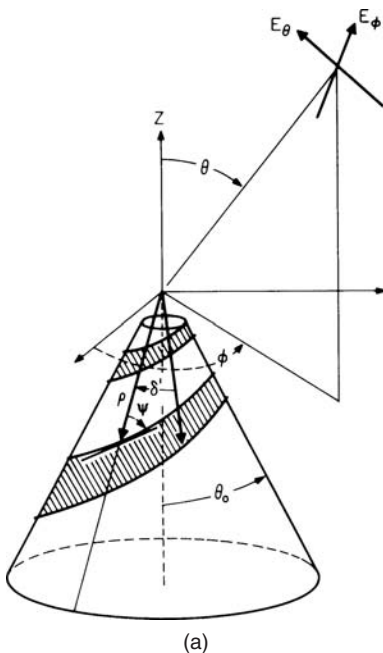


FIGURE 13-29 (a) Sketch of a conical spiral (after J. D. Dyson¹¹ © IEEE 1965) and (b) two four-arm spirals (Courtesy of GTE-Sylvania Systems Group)

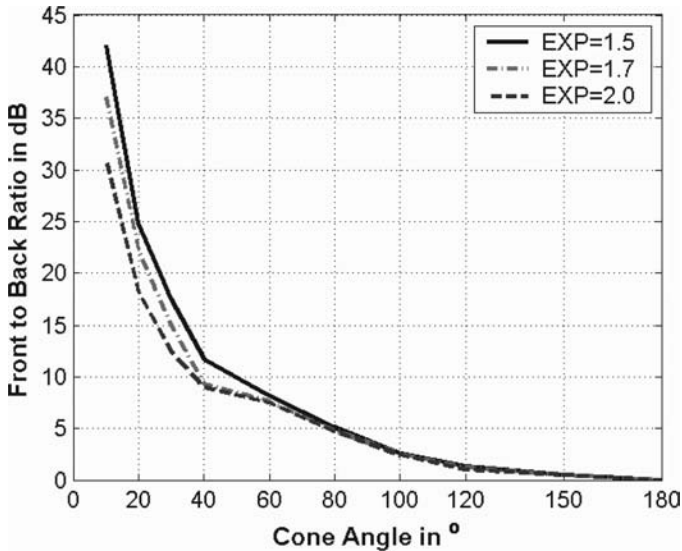


FIGURE 13-30 Conical spiral average front-to-back ratio versus cone angle for various growth rates

radiation characteristics do not greatly change. This finding was important for truncation of the spiral, thus simplifying fabrication while reducing the feed-coax length and the height of the cone. Computed front to back (F/B) ratios for various cone apex angles of a four-arm conical spiral are shown in Figure 13-30. As the cone angle is decreased (the spiral gets taller), the backlobe is reduced, and the directivity increases. Directivity also increases when the spiral is wrapped tighter since more turns cross the active region and coupling between arms is enhanced.

Conical spirals designed to operate in the mode 2 have been utilized for surveillance applications. It was shown that the cone apex angle and other parameters of the spiral can be adjusted to peak the beam at desired angles (see Figure 13-31). These four-arm spirals can be excited with a single 180° hybrid or other two-arm spiral feeding methods. Since at the feed opposite arms have the same excitation, they can be shorted together (without shorting to adjacent arms), leaving only 0° and 180° phasing requirements, the same as a two-arm spiral.

Spiral Miniaturization

A spiral is a fast leaky wave antenna that performs at its best when the traveling wave phase velocity is as high as possible (as an example, see Figure 13-19). Thus, slowing the wave on the spiral arms, though possible, must be exercised with extreme care. Techniques for slowing the wave on the spiral arms include loading the spiral with higher contrast/thickness dielectric material, increasing the length of the spiral arms using modulations or a modified spiral shape, and capacitive and inductive loading of the line. Though some techniques result in acceptable patterns and improved matching conditions, the antenna efficiency typically is reduced (for dielectric loading) and some pattern contamination may occur. A good example of extended low-frequency performance is given in *Four-Arm Spirals*,¹³ where a

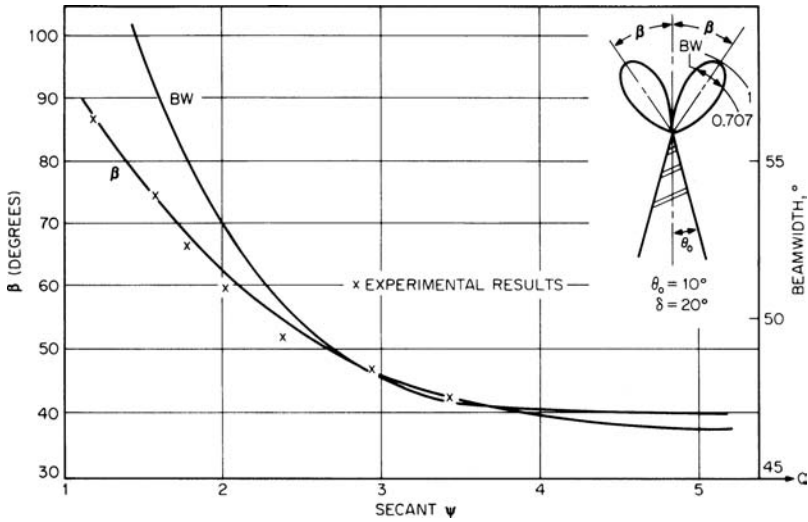


FIGURE 13-31 HPBW and beam-peak elevation angle for a 2nd mode four-arm conical spiral⁴

25-cm helix-loaded spiral demonstrated good patterns down to 250 MHz. A similar commercially available realization is shown in Figure 13-32a. As demonstrated by Filipovic and Volakis,³⁴ a 15-cm diameter two-arm meanderline slot spiral, shown in Figure 13-32b, showed good performance down to 250 MHz. The three-dimensional undulations and resulting inductive loading with the variable thickness dielectric (see Figure 13-33) are used by Chen and Volakis²² and impressive low-frequency gain values are demonstrated.

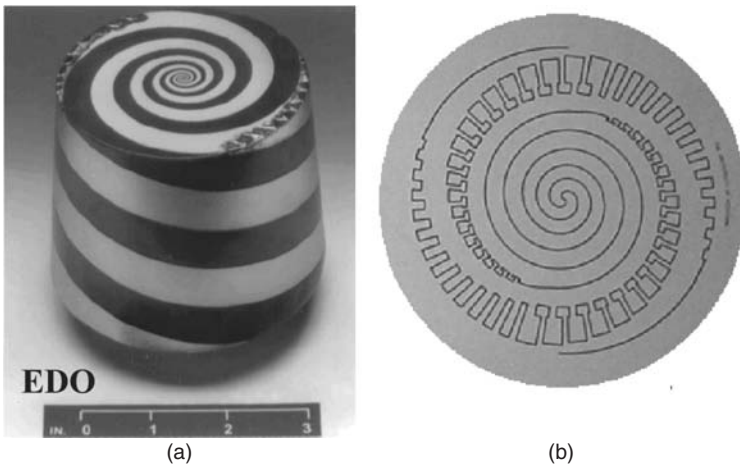


FIGURE 13-32 (a) Helix-loaded spiral (Courtesy of EDO) and (b) in-plane meandered slot spiral

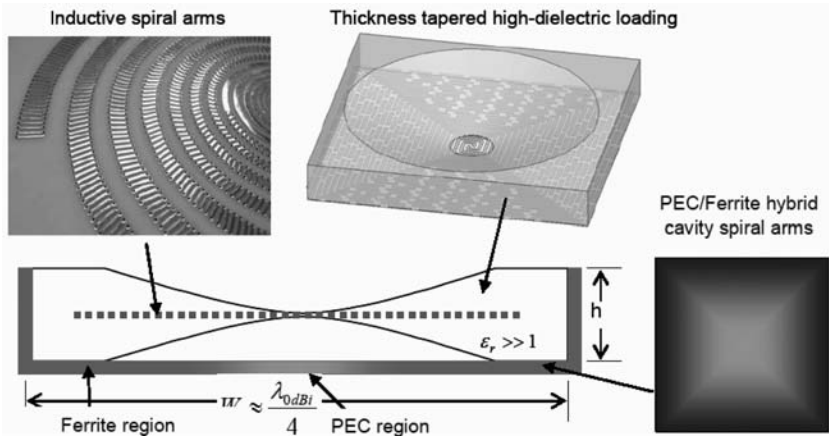


FIGURE 13-33 Spiral miniaturization by combining inductive and dielectric loadings (after C-C Chen and J. L. Volakis²² © Wiley InterScience 2007)

Beamforming

Beamformers, also known as *modeformers*, are used to provide proper phasing of the spiral arms to achieve the desired mode excitations. Phase excitations for the first two modes of four-arm RCP and LCP spirals are shown in Figure 13-34. For a counterclockwise (CCW)–wound RCP spiral, phase excitation of successive spiral arms is given by $p = \frac{-360m(n-1)}{N}$ (in degrees), where $m = 1, 2$ is the mode number, $n = 1, 2, 3, 4$ is the arm number, and $N = 4$ is the number of arms. Note that the theoretical phase distributions are identical for modes 2 RCP and LCP and that the spiral's wrap direction determines the polarization sense. The beamforming networks are typically stripline circuits built as a combination of 90° and 180° hybrids and phase shifters. An excellent treatment of spiral beamformers can be found in *Four-Arm Spirals*.¹³

The modal decomposition of the mode 1 input of a 4×4 Butler matrix beamformer is depicted in Figure 13-35. Complex S_{N1} data are measured from the mode 1 input to each of the four outputs of the beamformer ($N = 1, 2, 3, 4$). This data is then decomposed into the four spiral modes. Both amplitude and phase imbalances contribute to the modal contamination. The top curve is the insertion loss for mode 1, while the remaining curves (modal isolation > 25 dB) represent the contamination level of each of undesired modes, m_2 , m_3 , and m_4 .

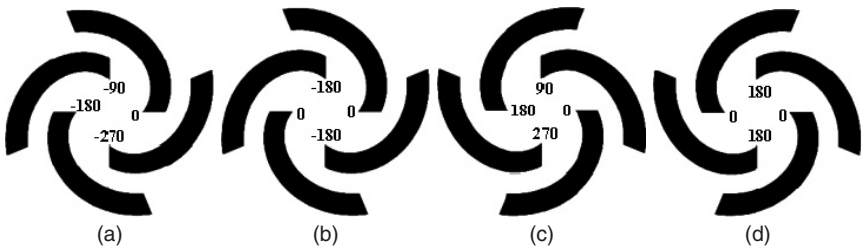


FIGURE 13-34 Four-arm phase excitations: (a) mode 1 RCP, (b) mode 2 RCP, (c) mode 1 LCP, and (d) mode 2 LCP

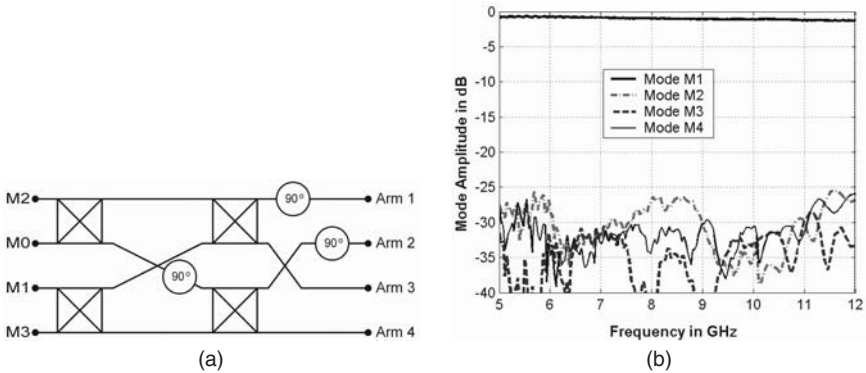


FIGURE 13-35 (a) Four-by-four Butler matrix beamformer and (b) modal decomposition for the spiral modes of a beamformer designed for 5–12 GHz operation

An ideal spiral integrated with this beamformer will have these modal pattern errors, assuming the spiral is large enough to support all the modes. Modal purity of other ports (m_2 , m_3 , and m_4) is determined in the same manner. Mode 4 will not radiate well from a four-arm spiral, but the beamformer modal contamination should be low so it will not excite direct radiation from the spiral feed structure. In general, for an N -arm spiral, mode N has the same input phasing as mode 0, and neither mode typically radiates efficiently. Mode 0 has no radiation region on the spiral, and the mode N impedance is extremely high as the equal phase currents in the feed region have difficulty propagating.

Reflector Feeds

Since spirals have approximately constant beamwidths they make ideal reflector feeds if broadband circular polarization is required. For planar absorptive cavity spirals, a reflector F/D of approximately 0.35 is optimal for mode 1. The spiral phase center is relatively stable across wide frequency ranges and is slightly in front of the spiral element, outside of the cavity, for this configuration. Figure 13-36 shows amplitude taper loss (ATL) and spillover

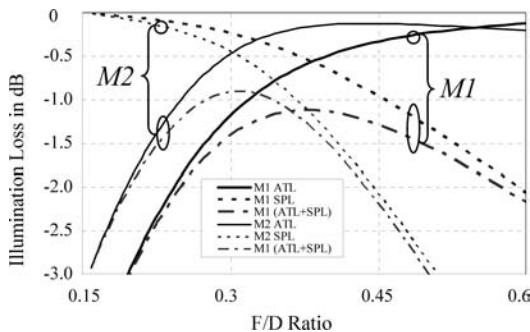


FIGURE 13-36 Aperture losses for a center-fed reflector with a four-arm planar spiral feed

loss (SPL) versus F/D , for both modes 1 and 2, using measured patterns of a four-arm spiral feeding a center-fed reflector. Maximum mode 1 aperture efficiency is when the F/D is 0.38. Mode 2 is most efficient at a 0.31 F/D . Simultaneous illumination is best at $F/D = 0.36$. Note the ATL + SPL curves are relatively flat and the optimal F/D may be chosen based on other factors such as maximum mode 1 gain or lower sidelobes. With a reflective cavity, the spiral directivity increases and the optimum F/D ratios increase to 0.45 for m_1 , 0.38 for m_2 , and 0.45 for simultaneous illumination of both modes.

Conical spirals have also been used as reflector feeds. The major advantage of conical spirals is improved efficiency due to backlobe reduction. A disadvantage is that the phase center moves with frequency resulting in a quadratic phase loss with frequency for a symmetric reflector. For this reason, the conical spiral feed can have optimal performance at only one frequency. Also, phase centers are not collocated for multiple modes at the same frequency. When used in an offset reflector, a conical feed will have beam-squint versus frequency due to phase center movement. Even with these limitations, conical spirals can be efficient feeds over broad bandwidths. Shown in Table 13-4 are illumination losses of less than 1.5 dB for 10:1 bandwidths using cone angles ($2\theta_0$) of 20 to 40 degrees. More extensive coverage is provided in Milligan's *Modern Antenna Design*.⁵

Lens Effects

Spirals are often realized with contact lenses^{5,13} made as a constant dielectric (hemi) spherical, Luneburg axicon, or as specially designed surfaces for custom-made dielectrics. Contact lenses can improve spiral efficiency by reducing the mismatch between the spiral aperture and free-space. The lower plane wave impedance in the lens medium allows for the spiral/lens configuration to radiate more energy into the upper hemisphere and less energy into the cavity. Care must be taken to ensure the lens substrate material near the spiral has very low loss. Since the spiral transmission region is generally very long (several wavelengths), and the spiral is loaded by this dielectric, losses quickly add up if the lens loss tangent is high. Hemispherical contact lenses will not dramatically change the beamwidth of a planar spiral. Other lens configurations can be made to focus the beam and substantially increase directivity and gain. Multiple spiral modes can be effectively focused using these lenses, and the angle-of-arrival (AOA) capabilities are retained (see Figure 13-37). Wide-angle lenses have been constructed for azimuth symmetric beams that steer out to 60° and beyond. Multimode spirals can be used as feeds and azimuth AOA can be obtained (elevation AOA is usually severely degraded for these types of lenses). Noncontact lenses can also be used, but the spiral efficiency will not be improved. Conical spirals are usually not used due to the short focal length of the lenses and the phase center movement of the conical antenna versus frequency.

TABLE 13-4 Optimum Conical Spiral Mode 1 Feeds for Paraboloid Reflector for 10:1 Bandwidth (after Milligan⁵ © J. Wiley & Sons 2005)

Cone Angle $2\theta_0, ^\circ$	f/D	Average Phase Error Loss, dB	Max Phase Error Loss, dB	Taper Loss, dB	Spill-over Loss, dB	Cross Pol. Loss, dB	Average Total, dB
20	0.46	0.36	1.00	0.43	0.54	0.09	1.42
30	0.42	0.22	0.60	0.42	0.63	0.16	1.45
40	0.38	0.19	0.55	0.50	0.60	0.28	1.56

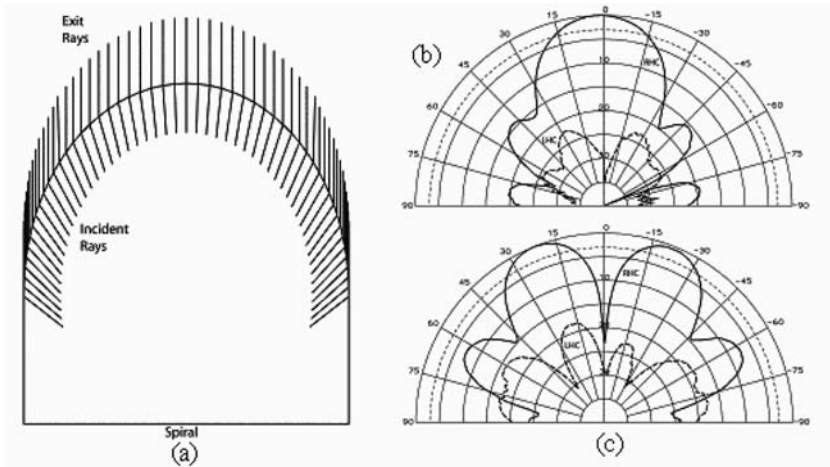


FIGURE 13-37 (a) Contact lens with ray trace, (b) fed with a mode 1 spiral, and (c) fed with a mode 1 spiral (after Milligan⁵ © J. Wiley & Sons 2005).

Other Spiral Geometries

Spiral antennas have been fabricated in a wide range of geometries not only for improved performance, but also for the reasons of packaging, arraying, and conformal and flush-mounting. For example, square spirals (see Figure 13-38) allow closer packaging in an array environment, and because of longer circumference than round spirals, provide some inherent miniaturization factor ($MF = 4/\pi$) for improved low-frequency operation. Wiggly spirals have modulations in the form of a zig-zag, sinusoid, or meander-line added to the spiral arms to increase arm length and also enhance low-frequency performance. These methods can improve low-frequency performance, but overmoding is still present at higher frequencies. Spirals have also been made on parabolic reflector surfaces and conformal to spherical bodies.

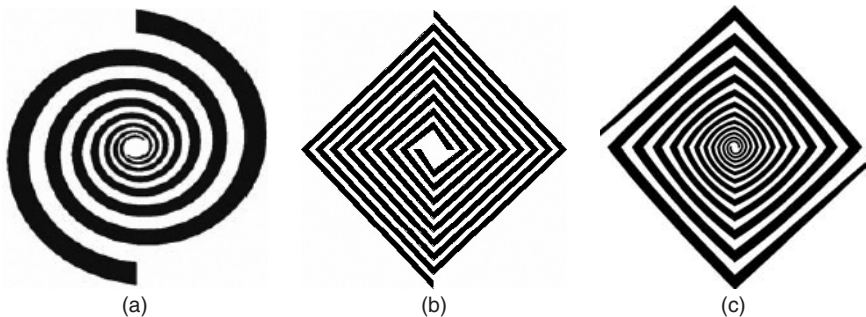


FIGURE 13-38 Circular, square, and hybrid spiral lay-outs

Spiral Modeling

The complexity of spiral geometry and absence of appropriate computational tools (including methods for numerical electromagnetics) in the 1950s limited the study of spiral antennas to primarily experimental efforts. The first theoretical analysis of a spiral antenna was published in 1960,³⁵ where the spiral geometry was approximated by semicircles, and the vector potentials were computed by employing the appropriate Green's function. Although this was a simplified model, the physical understanding of the spiral's operation, including the low- and high-frequency limits, were obtained from this analysis. The availability of faster computers and the introduction of rigorous analysis methods for solving integral equations (numerical methods) was critical for comprehensive theoretical investigation of spiral antennas. Nevertheless, while integral equation methods have been traditionally used for free-standing printed and wire spirals, and for noncavity-backed slot spirals, partial differential equation and volume modeling methods are more appropriate for the cavity-backed spirals and spirals on a finite substrate. The finite element-boundary integral method (FE-BI), developed specially for the cavity-backed Archimedean slot spirals,³⁶ has produced very accurate results.^{23,34} This method combines triangular and quadrilateral prismatic finite elements to obtain the best representation of the electric fields inside an extremely narrow slotline.

Recently, a finite volume-time domain technique (FVTD)³⁷ was used to accurately model realistic spiral geometry, as shown in Figure 13-39. Complex, honeycomb absorber loading, balun, thin dielectric, and fine details of a tightly wound Archimedean spiral are modeled for the first time in its entirety and excellent agreement with measurements has been obtained (see Figure 13-40). Finally, the analysis of spiral antennas has also been performed using commercially available numerical tools including method of moments codes, NEC, FEKO, Ansoft Designer, IE3D, as well as finite element software tools such as Ansoft HFSS and finite difference time domain XFDTD. A comparison between the HFSS, FEKO, and measurements, shown in Figure 13-41, clearly demonstrates the accuracy of these tools in modeling spiral and other FI antennas.

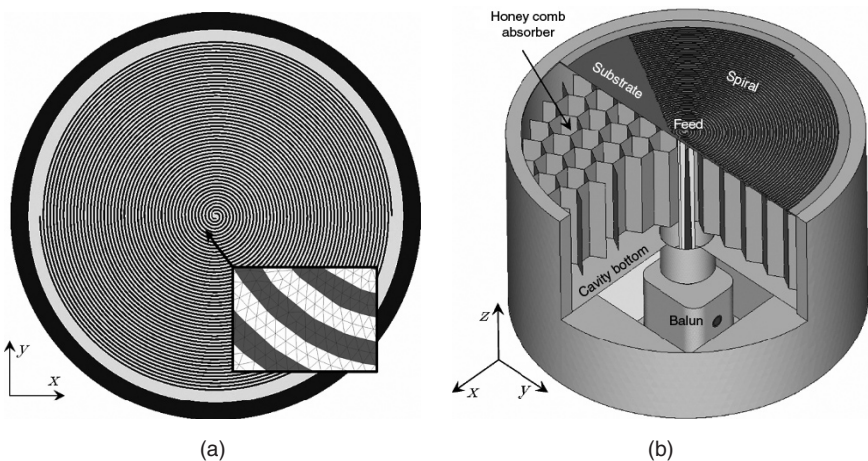


FIGURE 13-39 Geometry of a tightly wound spiral over a realistic absorber filled cavity modeled using the FVTD technique (after C. Fumeaux et al³⁷ © IEEE 2006)

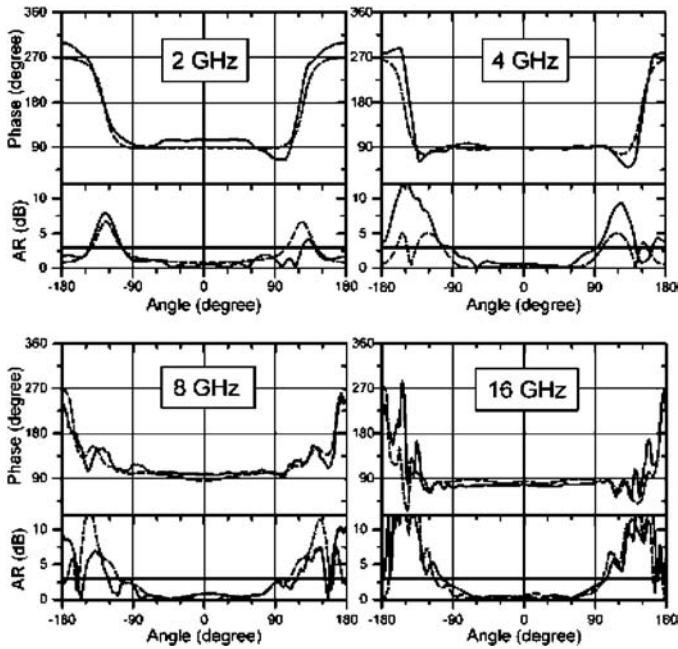


FIGURE 13-40 Axial ratio of the spiral shown in Figure 13-34: comparison between measurements and simulations using the FVTD technique (after C. Fumeaux et al.³⁷ © IEEE 2006)

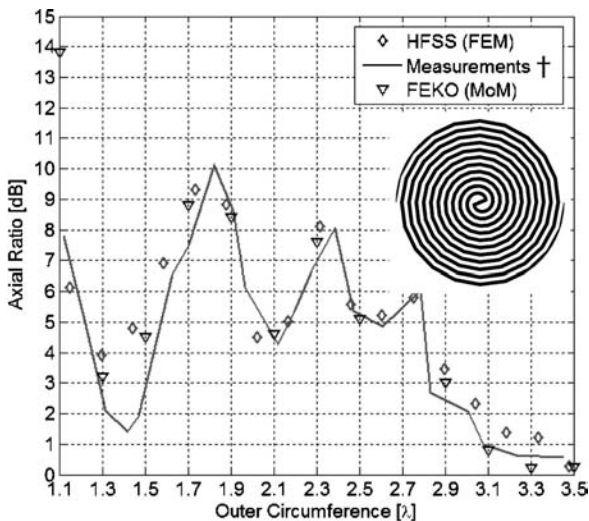


FIGURE 13-41 Comparison between two commercial software tools, Ansoft HFSS and EMSys FEKO, with measured results from “A spiral antenna backed by a conducting plane reflector” by H. Nakano, K. Nogami, H. Mimaki, and J. Yamauchi³⁸ (after M. Buck and D. Filipovic³² © IEEE 2006)

13.3 LOG-PERIODIC ANTENNAS

Log-periodic antennas were introduced in the late 1950s by DuHammel and Isbell.⁸ They derived the analytical formulation of log-periodic principles, including the mathematical description of the geometry, and experimentally verified the proposed principles on a planar sheet (slot) antenna, as shown in Figure 13-4. Continuing this original work and the unpublished work of Isbell, DuHammel and Ore³⁹ noticed that the currents were concentrating on the edges of a planar self-complementary metallic sheet, so they removed the metal from the center and wire-like log-periodic antennas were born. They also demonstrated circularly polarized designs and unidirectional performance with a folded geometry. Design guidelines and the dipole log-periodic antenna with transposed feed, one of the most frequently used radiators, were established by Carrel.⁴⁰ A significant number of contributions have been reported since these early days,^{3,4,5} however, we will only mention the important discovery of a frequency independent log-periodic sinuous antenna by DuHammel.⁴¹ At the time, this was the first single-feed, planar antenna element that was simultaneously multi-octave wideband and omnipolarized.⁴²

Basic Principles

A *log-periodic antenna* is defined as a structure whose electrical properties vary periodically with the logarithm of frequency. If an antenna is realized such that within each period its performance changes within bandwidth-defined conditions (for example, the real part of the input impedance = $70 \pm 20\Omega$, or the gain = 5 ± 1 dBi, etc.), then multi-decade operation can be obtained. To produce the described performance, similar mathematical relationships must be applied to the antenna's structural parameters. This is well depicted in Figure 13-2, where the two arms of a six-arm log-periodic antenna is shown. A log-periodic structure can be obtained by taking an appropriate 2D cut of a 3D FI antenna known as a conical screw.⁷ As shown in Figure 13-42, the planar log-periodic antenna is nothing else but a constant ϕ -cut of a conical screw. Following the discussion

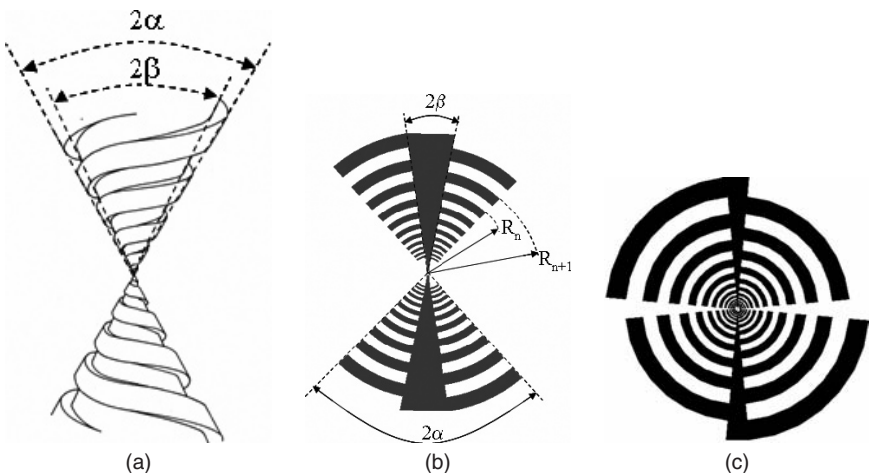


FIGURE 13-42 (a) Conical screw, (b) planar log-periodic antenna obtained from the conical screw as $\phi = \text{const}$, and (c) self-complementary two-arm structure with curved teeth

from Section 13.1, it is clear that this structure is not continuously self-scaled with the frequency. However, for a discrete set of frequencies, such that $\log(f_n/f_{n-1}) = \text{const}$, the structure and the bandwidth scale the same way.

Assuming an infinite log-periodic geometry with an infinitesimally precise and fine center region, one observes that the current distributions at two frequencies separated by the ratio $\ln(1/\tau)$, where $\tau = R_n/R_{n+1}$, is the same. This gives rise to the same input impedance with the same period, while the period of radiation pattern is $2\ln(1/\tau)$.⁴ The computed input impedance of a nonplanar log-periodic antenna with 26 elements and a growth rate of $\tau = 0.85$ is shown in Figure 13-43. The variation of the period (as depicted inside the plot) is a consequence of finite antenna size and imprecision in reading out the x -axis values. However, the previously mentioned consequence of the log-periodic growth is clearly demonstrated. As with the spiral antennas, the overall bandwidth is limited by the fine precision of the center region (high end) and finite size (lower end) of the log-periodic radiator.

In this section, our overview includes planar bidirectional and unidirectional log-periodic tooth radiators, as well as end-fire log-periodic dipole and slot arrays. The basic principles of operation, typical performance including multimoding, feeding, and application as reflector feeds are also discussed. For additional reading, see the literature.^{2,4,5}

Planar Log-Periodic Tooth Antennas

Experimental verification of a decade-bandwidth log-periodic FI design was first demonstrated in DuHamel and Isbell's paper, "Broadband Logarithmically Periodic Antenna Structures."⁸ The planar free-standing element, shown in Figure 13-4, is a linearly polarized bidirectional radiator. Very efficient attenuation through the active region composed of several teeth in the neighborhood of a half-wavelength long pair, results in little (if any) observable end effect. The teeth can be of circular or straight/trapezoidal shape with respective antennas showing little difference in performance.³⁹ For bidirectional operation, the

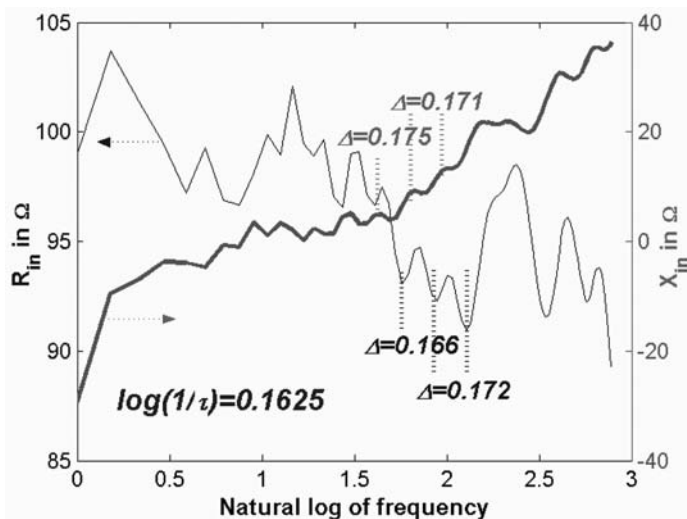


FIGURE 13-43 A consequence of a log-periodic design: depicted is the constant resistance and reactance period when plotted vs. $\ln(f)$.

feed is planar and is realized using a coaxial cable soldered to the metal portion between the teeth (equivalent to the Dyson balun). The coax transitions to the slot in the geometrical center of the antenna aperture. When unidirectional operation of a planar structure backed by a cavity is sought, a vertical balun feed is typically utilized (and also serves as an impedance transformer). The typical impedance at the balanced ports of an antenna with self-complementary geometry is $\sim 160\Omega$. This is slightly lower than the theoretical value of 188.5Ω due to the non-zero metallization thickness, finite features of the feed region, dielectric support (if used), and so on. Lowering the impedance by increasing the metal to slot ratio (deviation from the self-complementary structure), has proven to be very challenging since impedance and pattern variations, both within the period and the bandwidth, are typically significant.

Printed log-periodic antennas are typically designed with two, four, or eight arms. A two-arm planar antenna is linearly polarized, while a four-arm geometry with the phase progression of $\pm 90^\circ$ between the orthogonal pairs of arms, usually implemented with a 90-degree hybrid, is circularly polarized (mode ± 1). Due to the phase ambiguity associated with the excitation of the higher-order modes, more than four arms are needed when circularly polarized multimode operation is desired. The simplicity of the BFN is the main reason why eight-arm geometries are commonly utilized when multimode operation is required. As mentioned, planar sheet log-periodic antennas are typically designed with curved or straight teeth,³⁹ with performance between the two being very similar. The main difference is in the variation of the half-power beamwidth over the operating band. Shown in Figure 13-44 are the averaged beamwidths of four-arm structures when fed to operate in the circularly and linearly polarized modes. The two antennas are self-complementary with the same diameter of 7.5 cm, a growth rate of $\tau = 0.8$, a boom width of $2\beta = 10^\circ$, a tooth swing of $2\alpha = 80^\circ$, and a number of teeth $N = 17$. Consistently, the curved log-periodic has a lesser variation of the 3 dB beamwidth over the wide 9:1 bandwidth. It is important to note that the low-frequency gain and impedance match of the antenna with the straight tooth is slightly better. This is due to the longer length of the teeth and associated square vs. circular antenna aperture. Interestingly, the self-complementary two-arm planar sheet antennas (see Figure 13-42c) typically have several narrow-band VSWR, gain/pattern drop-outs. This can be resolved with the smaller teeth swings and subsequent deviation from the self-complementary principle.

The performance of planar log-periodic antennas is affected by the growth rate $\tau (= R_{n+1}/R_n)$, boom angle 2β , also referred to as the angular separation between the teeth, the angular arm width or swing 2α determining the interleaving between the neighboring arms, and dielectric

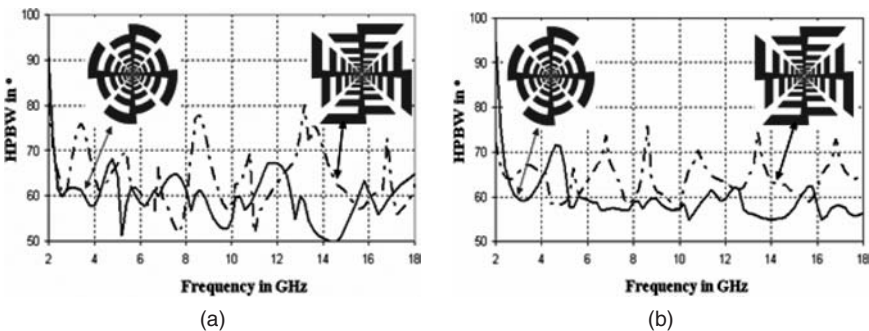


FIGURE 13-44 Average 3 dB beamwidth for four-arm planar log-periodic antennas with curved and straight teeth shown in insets when they are fed to operate in circularly (a) or linearly (b) polarized mode

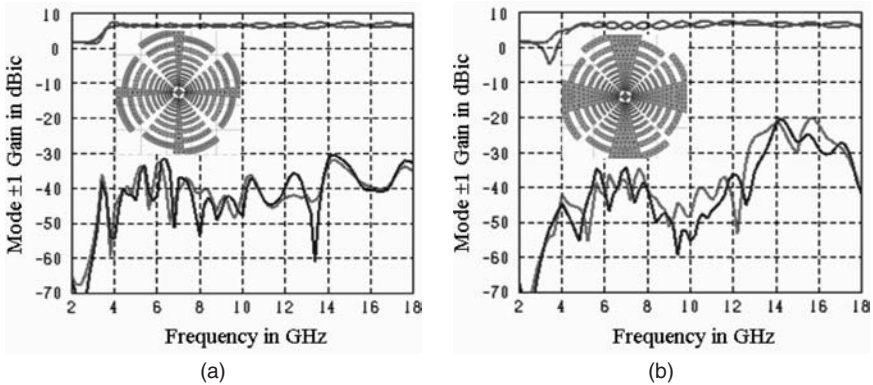


FIGURE 13-45 Computed effects of the boom angle 2β on the broadside co- and cross-polarized mode ± 1 gains of 6-cm diameter four-arm planar log-periodic antennas shown in the inset. Boom angles are $2\beta = 10^\circ$ (a) and $2\beta = 30^\circ$ (b).

constant of the substrate. The structural parameters are as depicted in Figure 13-42. Effects of the boom angle on the performance of a free-standing four-arm planar log-periodic antenna with a diameter of 7.5 cm, $N = 22$ elements, $\tau = 0.82$, and $2\alpha = 80^\circ$ are computed with FEKO, and the results are shown in Figure 13-45. The antennas are fed for circularly polarized modes ± 1 operation. Clearly, an antenna with a narrower boom angle has improved low- and high-frequency operation. Specifically, for the same aperture size, the antenna low-frequency gain, and high-frequency cross-polarization (axial ratio) both improve with the reduced angle 2β . This can be anticipated since the width of the boom affects the length of the teeth and thus the resonant frequency of the lowest active region.

Several planar, printed eight-arm log-periodic layouts, with different lengths of circular teeth (2α) are shown in Figure 13-46. With an appropriately designed mode-forming network, these antennas can produce high-quality multimode and omnipolarized performance with stable modal impedances, consistent patterns, and cross-polarization levels. Typical performance of free-standing structures with an aperture diameter of 6 cm, $\tau = 0.73$, $2\beta = 5^\circ$, and $2\alpha = 40^\circ$ – 70° , for several modes, is shown in Table 13-5. Note that the nominal impedance decreases with the increased mode number, similarly as with multimode spirals. They can be computed using Deschamps's formula (Eq. 13-11). Also, with a more pronounced interleaving between the neighboring arms ($2\alpha \uparrow$), the variation

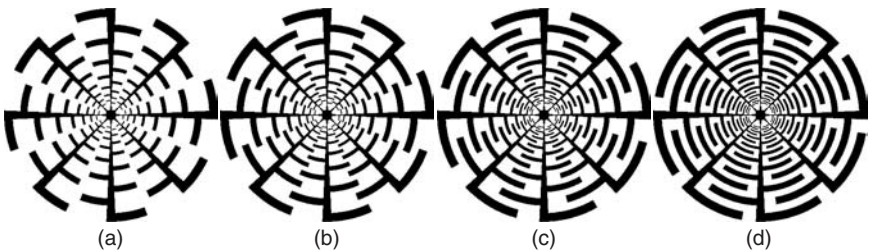


FIGURE 13-46 Various planar log-periodic eight-arm structures with the arm widths (2α) of 40° (a), 50° (b), 60° (c), and 70° (d)

TABLE 13-5 Averaged Characteristics of Multimode Freestanding Planar, Log-periodic, Eight-arm Structures Shown in Figure 13-46

2α [°]	Mode 1			Mode 2			Mode 3		
	R [Ω]	G [dB]	AR [dB]	R [Ω]	G [dB]	AR [dB]	R [Ω]	G [dB]	AR [dB]
40	300	6.6	0.1	177	4.5	1.0	126	3.1	3.9
50	294	6.6	0.1	158	4.0	1.1	118	2.4	4.0
60	285	6.3	0.1	154	3.4	1.8	109	1.5	6.2
70	285	5.7	0.1	131	2.6	2.0	91	0.3	17

of both, the resistance, and reactance over the operating band increases. As with other multimode FI antennas, the nominal gain and axial ratio at the elevation of the maximum gain both decrease with increased mode numbers. Increased interleaving deteriorates these two far-field parameters, to the point that structures with $2\alpha \geq 60^\circ$ have several frequency bands where they are linearly polarized when fed for mode 3 operation. For example, for $2\alpha = 70^\circ$, the axial ratio of mode 3 is always greater than 4 dB within the range of this mode's operation.

The increased tooth interleaving contributes to the movement and more pronounced confinement of the radiating region toward the inside of the log-periodic aperture. As a consequence, the radiation from the higher-order modes becomes noticeable via the deteriorated symmetry of the radiation patterns. Shown in Table 13-6 are selected far-field characteristics at 12 GHz for all three copolarized broadband modes. A larger 2α contributes to the symmetric dislocation, or squint, of a mode 1 beam maximum from broadside. The position of the mode 2 and 3 beam maxima remains stable with small and moderate interleaving. The WoW, in general, deteriorates with the increased 2α ; however, for large interleaving, the pattern becomes highly contaminated, as shown in Figure 13-47. Notice the gain reduction due to the miniaturization effects. For multimode operation, these antennas are fed the same way as spirals when a single polarization is needed or as sinuous antennas when dual-polarized operation is required.

Finally, an unusual, yet for some applications desired, performance can be obtained if ambiguous feeding is used. For example, for a four-arm antenna, phase progression of 0° ,

TABLE 13-6 Far-field Radiation Pattern Characteristics at 12 GHz: The Elevation Angle of Maximum Copolarized Gain, WoW at That Angle, and WoW at 60° Computed for Modes M1, M2, and M3

2α [°]	$\theta_{-G_{\max}}$ [°]			WoW at $\theta_{-G_{\max}}$ [dB]			WoW at $\theta = 60^\circ$ [dB]		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
40	0	31	41	0.00	0.33	0.77	1.10	0.53	1.79
50	3	33	42	0.03	0.40	1.15	1.17	1.08	2.24
60	7	33	43	0.06	0.38	1.16	0.75	1.00	2.50
70	23	25	30	0.14	0.37	0.64	0.43	1.80	6.70

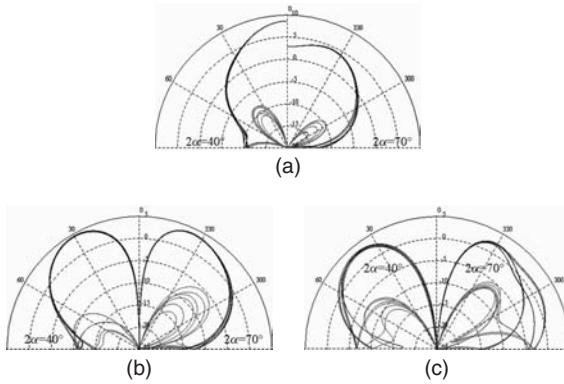


FIGURE 13-47 Modes 1 (a), 2 (b), and 3 (c) radiation patterns for eight-arm planar log-periodic antennas from Figure 13-5. Shown are six overlaid co- and cross-polarized patterns taken with 30° separation in azimuth.

180° , 0° , 180° can produce equally well left- and right-handed circular polarization since this antenna aperture does not have polarization filtering capability as spirals (wrapping direction). As a consequence, the four-arm printed log-periodic antenna radiates a monopole-like horizontally polarized pattern along the symmetry planes of the arms and a vertically polarized pattern along the planes between the arms. The total gain pattern is fairly omnidirectional. Shown in Figure 13-48a and b are computed three-dimensional linear gain patterns at the midband frequency obtained with FEKO. Also shown in Figure 13-48c are overlaid co- and cross-polarization patterns at six discrete frequencies over a 4:1 band in two constant azimuthal cuts and in Figure 13-48d, a constant elevation cut at $\theta = 38^\circ$. The radiation patterns are very consistent over a wide frequency range, with the polarization sense changing every $\phi = 45^\circ$. A deep null at broadside, with max gains around $\theta = 38^\circ$, good cross-polarization, and matched gain values close to the gain of a half-wave dipole are the main properties of this antenna. This unconventional frequency independent log-periodic antenna is well matched throughout the operating band.

Non-Planar Log-Periodic Tooth Antennas

Efficient unidirectional operation can be achieved by folding the planar arms, i.e., by reducing the angle between the axes of the arms⁴ or by projecting the planar log-periodic structure on a cone.⁴³ The latter is more complicated to design, and since they perform similarly, the former will not be described further. The angle between the folded arms is denoted ψ in Figure 13-49. For a planar configuration $\psi = 180^\circ$, note that the teeth are straight, with the relative width computed as $\varepsilon = r_N/R_N$ (typically $\varepsilon > \tau$). Folding eliminates the need for lossless reflective (band-limited) and/or lossy absorptive (broadband) cavities and typically yields about a 2–2.5 dB increase in forward gain when compared to the free-standing (or absorptive cavity-backed) antenna. Additionally, smoother frequency response can be obtained with fewer teeth, and simpler feeding can be designed. To reduce the antenna weight and windload, the solid metallic arms of a free-standing structure can be replaced by a thin metallic wire or strip that outlines the arms.³⁹ Changes in impedance and pattern are minimal since the current is concentrated along the edges of the original geometry.

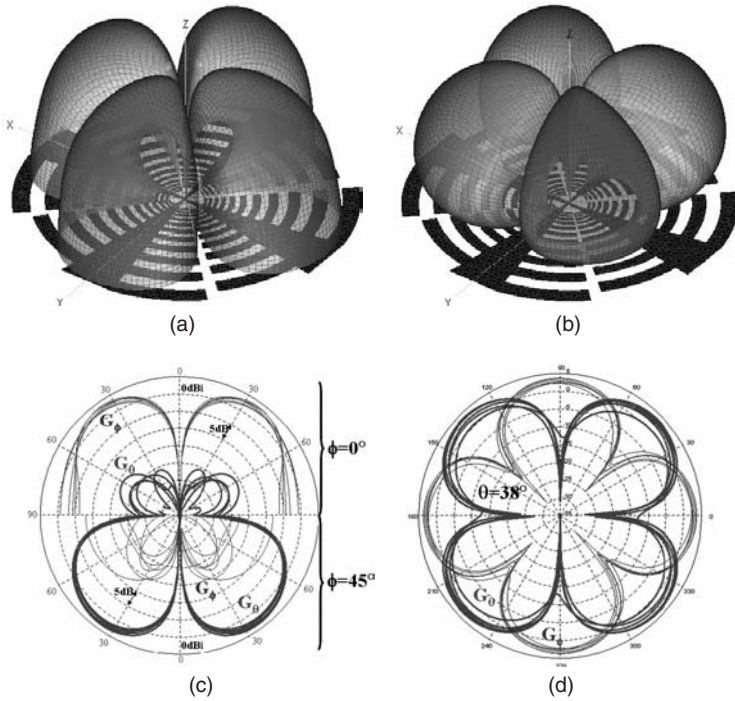


FIGURE 13-48 Computed 3D gains G_ϕ (a) and G_θ (b) at midband for a four-arm log-periodic planar tooth antenna with overall phase progression between the arms of 720° . Also shown are constant azimuth (c) and constant elevation (d) cut co- and cross-polarization patterns for six discrete frequencies over a 4:1 frequency range.

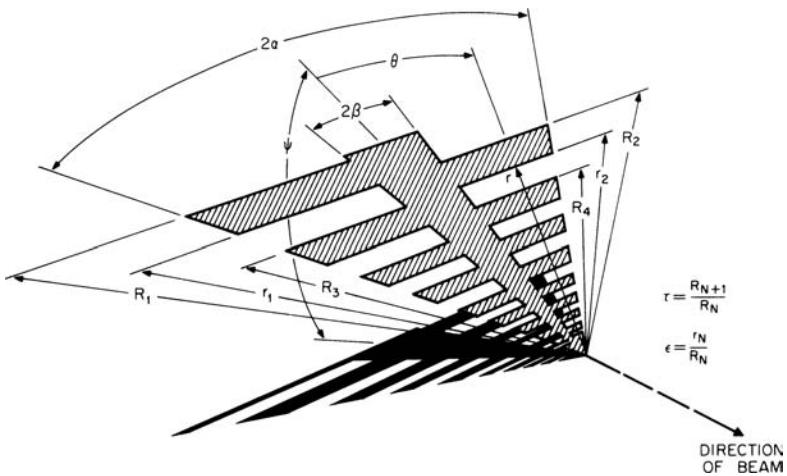


FIGURE 13-49 Trapezoidal-tooth log-periodic antenna (after R. H. DuHamel and F. R. Ore³⁹ © IEEE 1958)

Due to the small variations of antenna impedance and pattern over wide frequency ranges, these antennas are commonly used as wideband feeds for reflectors.⁴⁴ They can be fed with an infinite balun, thus eliminating the need for the crisscross feeders and significantly simplifying their construction. The only major disadvantage is the phase center movement versus frequency. A NEC wire model of a 1–18-GHz design is shown in Figure 13-50a. A total of 26 resonant dipole pairs (elements) are distributed along the boom with the growth rate of $\tau = 0.85$. Arms are subtended over an angle of $\psi = 5.6^\circ$. The boom width is $2\beta = 2.2^\circ$, while the arm width is $2\alpha = 33.4^\circ$. The feed is at the high frequency end, with a distance between the arms of 1.5 mm.

The nominal input resistance of this antenna is about 95Ω , and it can be tuned by varying the distance between the arms at the feed location. Shown in Figure 13-50b is the computed VSWR with respect to 50Ω and the antenna nominal impedance. When a broadband impedance transformer is used, this antenna can be matched for the $\text{VSWR} < 1.5:1$ throughout the designed frequency range. A direct 50Ω feed will still produce an adequate match with $\text{VSWR} < 2.3:1$ throughout the operating band.

The antenna HPBW and gain are shown in Figure 13-51. The E -plane variation throughout the 18:1 frequency range is about $\pm 5^\circ$, while the H -plane beamwidth changes about $\pm 14^\circ$ around the averaged value. Notice that the period varies with the logarithm of frequency. For many applications, the 10-dB beamwidth is important, particularly when used as the feed for a reflector antenna. The log-periodic antenna discussed in this paragraph has 10-dB E -plane and H -plane beamwidths in the ranges 99° – 108° and 130° – 203° , respectively.

It is interesting to note that the wider H -plane beam (and associated deviation around the nominal value) can be tuned by changing the angle ψ . By increasing the angular separation between the arms, the beam becomes narrower in the H -plane. At the same time, the E -plane beamwidth is only slightly changed. Thus, if needed, an antenna with very similar patterns in the E - and H -planes can be obtained. However, the drawback is reduced front-to-back ratio. Note also that directivity is increased due to the smaller H -plane beamwidth, and it can be further enhanced by increasing the growth rate τ and optimizing angle ψ . The nominal values for the 3-dB beamwidths in E - and H -planes as a function of ψ are shown in Table 13-7.

The nominal impedance of non-planar log-periodic tooth antennas can be tuned by changing the distance between the arms. Reduced spacing increases the capacitance of the

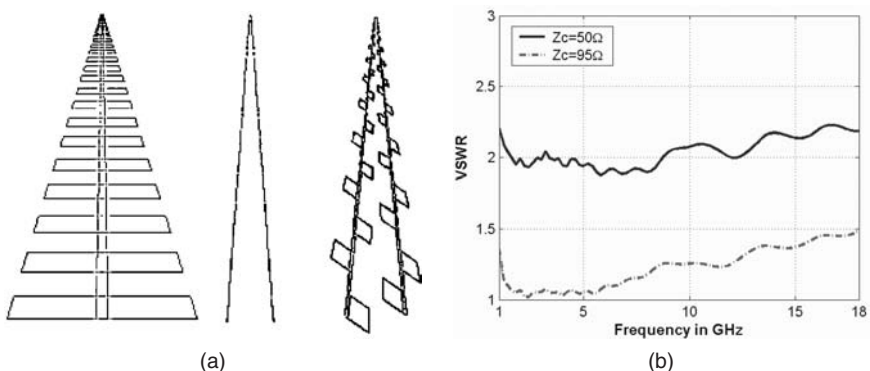


FIGURE 13-50 Sketch of a 26 element 1–18-GHz non-planar log-periodic tooth antenna with the following parameters: (a) $\psi = 5.6^\circ$, $2\alpha = 33.4^\circ$, $2\beta = 2.2^\circ$, $\tau = 0.85$, and (b) computed VSWR for 50 and 95Ω match

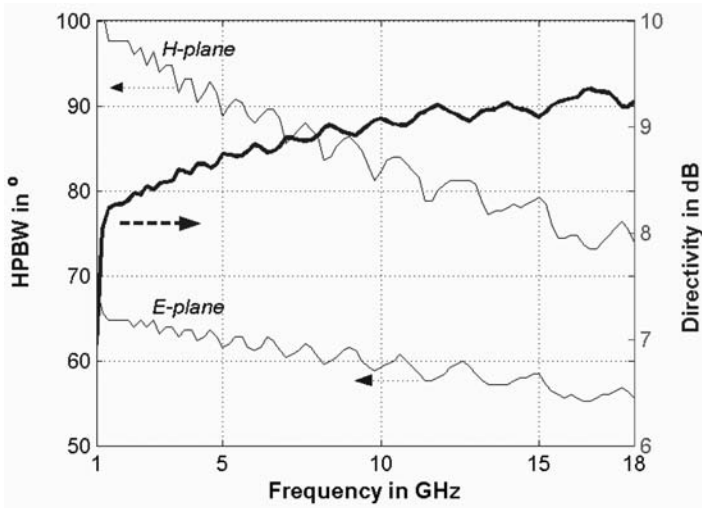


FIGURE 13-51 Computed *E*- and *H*-plane HPBW and gain versus frequency for a log-periodic antenna with $\psi = 5.6^\circ$, $2\alpha = 33.4^\circ$, $2\beta = 2.2^\circ$, and $\tau = 0.85$

transmission line, and thus its characteristic impedance. When the angle between the arms is maintained, the far-field is not significantly changed. The variation of input resistance and reactance for various spacings between arms is shown in Figure 13-52. The antenna parameters are $N = 26$, $\tau = 0.85$, $2\beta = 2.2^\circ$, $2\alpha = 33.4^\circ$, and $\Psi = 5.6^\circ$.

Log-Periodic Dipole Antennas

The log-periodic dipole array (LPDA) antenna was the first FI antenna that immediately after conception made a huge impact on both commercial and military applications throughout the world. Although Carrel's studies have provided a theoretical and experimental foundation for log-periodic dipole antennas,⁴⁰ the concept was first proposed by Isbell,⁴⁵

TABLE 13-7 Averaged Characteristics of a 1-18 GHz Non-planar Log-periodic Tooth Antenna for Various Angles Between the Arms (For this 26 element, log-periodic antenna $2\alpha = 33.4^\circ$, $\tau = 0.85$, and $2\beta = 2.2^\circ$)

$\psi (^\circ)$	0.9	2.6	5.5	8.5	11.4	14.4	17.5
HPBW <i>E</i> -plane($^\circ$)	60.0	60.0	60.2	60.7	61.4	62.3	59.1
Δ HPBW <i>E</i> -plane ($^\circ$)	11.6	10.8	10.4	10.8	14.4	16.4	17.2
HPBW <i>H</i> -plane ($^\circ$)	93.9	90.7	84.9	78.7	72.2	65.8	63.5
Δ HPBW <i>H</i> -plane ($^\circ$)	33.2	31.2	28.0	22.4	18.8	18.0	18.8
Directivity (dB)	8.4	8.6	8.9	9.2	9.4	9.5	9.5
Front-to-back ratio (dB)	16.1	16.7	16.6	15.3	13.6	11.4	9.2

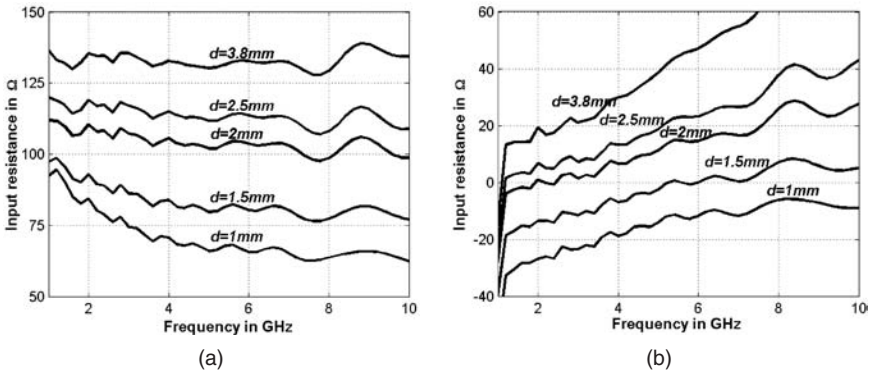


FIGURE 13-52 Computed impute resistance (a) and reactance (b) versus frequency for a log-periodic antenna with $\psi = 5.6^\circ$, $2\alpha = 33.4^\circ$, $2\beta = 2.2^\circ$, and $\tau = 0.85$. The variable is the distance between the booms (arms).

who made the following modifications to the structure he previously conceived with DuHamel:

- Arms of the tooth structure are folded, and their axes are made parallel. This corresponds to the limiting case of the angle $\psi \sim 0^\circ$.
- The width of the teeth $R_N - r_N$ is kept constant throughout the antenna structure.
- Planar circular teeth are replaced by straight wires with uniform cross-section.
- The apex angle of an embedded cone (in the basic conical screw geometry that, as discussed earlier, is the foundation for the planar sheet) approaches the limit 0° . In other words, the supporting booms are straight rods typically with rectangular cross-sections.

As a result of these modifications, a log-periodic wire dipole antenna emerged, most frequently used worldwide as a receiving antenna for terrestrial VHF/UHF radio and television (see Figure 13-53). The antenna consists of two in-plane sets of crisscrossed monopoles whose lengths and separations are related via the growth rate τ . The alternating feeding

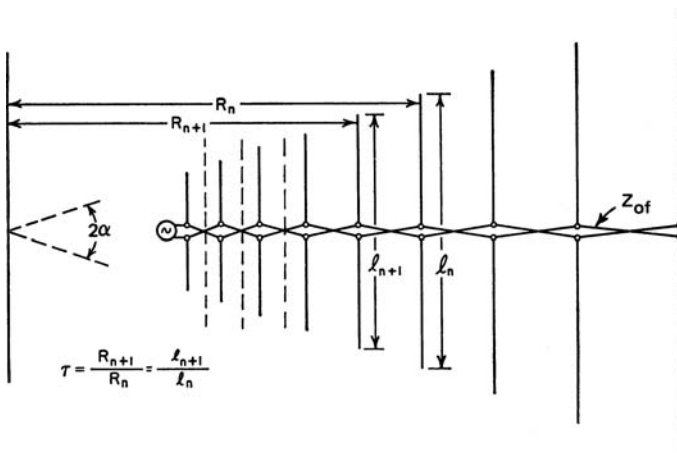


FIGURE 13-53 Schematic diagram of a log-periodic antenna. The three dashed lines indicate typical locations of electric ground planes between each pair of dipoles.

is required to achieve the backward end-fire radiation in the direction of the small dipole pairs. The currents on the consecutive monopoles in the transmission line region are in the opposite direction, thus contributing little to the overall radiation. However, the capacitive loading of open-ended monopoles produces the slow-wave effects in this region (typically $v = 0.6c$). The active region width depends on the antenna structural parameters (τ , α), and it occurs where the monopoles are about $\lambda/4$ long. With a properly designed antenna, the radiation losses in this region are large (20 dB or higher), so the residual currents are small and can be neglected. If the monopoles on each boom are directly connected such that the two-wire line is the feeder for the log-periodic dipole pairs, radiation would occur in the forward end-fire direction. It is clear that the effects of the non-resonant, capacitive dipoles in the transmission line region and the dipoles past the active region will be such that the far-field contamination would be significant. A typical performance of a well-designed log-periodic dipole antenna with $\tau = 0.95$ is characterized in Figure 13-54.

A systematic design of log-periodic dipole antennas was also provided by Carrel, and with some modifications can be found in many antenna textbooks.^{5,6} It is important to note that the scaling principle does not need to be applied to all structural parameters of this antenna, i.e., the boom tube diameter and the boom spacing can and usually are kept constant. Their effect on the antenna performance is minimal. To describe the structure, and eventually design the log-periodic dipole antenna, we are presented with two out of three important structural parameters: growth rate τ , angle 2α , and spacing factor σ defined as the normalized distance between a dipole and its closest, smaller dipole. The normalization is performed with respect to the length of the longer dipole. Along with the desired bandwidth, any two of these parameters are needed to describe the antenna structure. The spacing factor can be computed as

$$\sigma = 0.25(1 - \tau) \cot \alpha \quad (13-15)$$

The constant directivity counters, as originally given by Carrel⁴⁰ but corrected by Butson and Thompson,⁴⁶ as functions of τ and σ , are plotted in Figure 13-55. The feeder impedance of 100Ω and the ratio of monopole length to the width of 125 are taken as constants (see the literature^{47,48} for more extensive study).

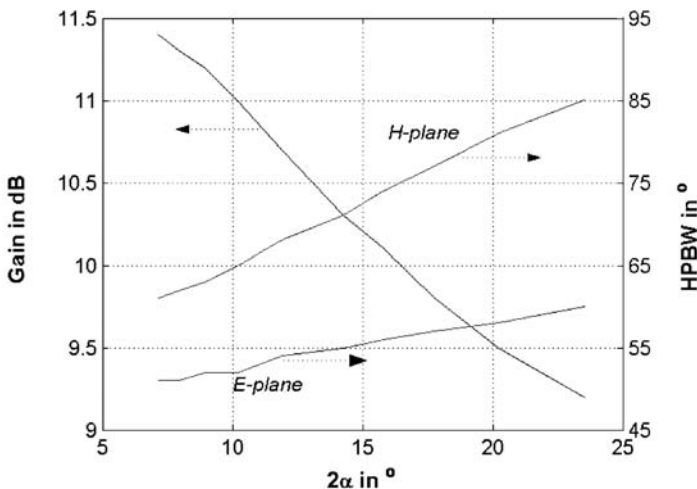


FIGURE 13-54 Computed average gain and 3-dB beamwidth in E- and H-planes of a log-periodic dipole array with $\tau = 0.95$

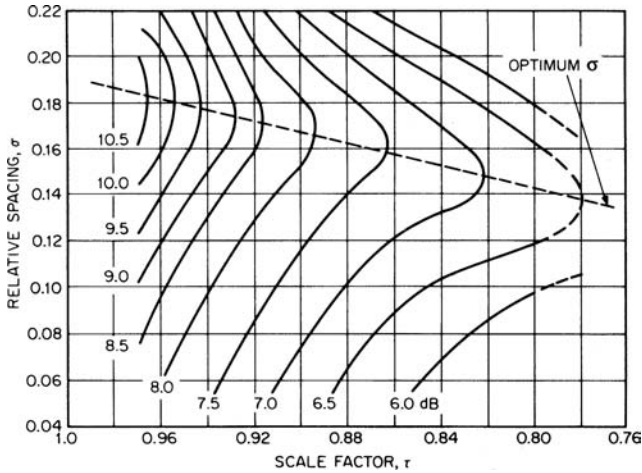


FIGURE 13-55 Constant-directivity contours in decibels versus τ and σ . Optimum σ indicates maximum directivity for a given value of τ (after R. L. Carrel⁴⁰ © IRE 1961).

As discussed, the log-periodic dipole antenna is fed at the high-frequency end, thus achieving end-fire radiation along the direction of the smaller monopole lengths. A two-wire transmission line can be used for direct excitation of the monopoles placed at the booms. However, a much simpler method of feeding is to use a coaxial cable inside one of the hollow booms, and at the high-frequency end attach its outer conductor to the host boom and the inner conductor to the opposite boom tube. This is a broadband balanced feed and to some extent resembles an infinite balun. When used for terrestrial television, the coax is usually 75Ω and a 4:1 impedance transformer is used at the TV set to convert to 300Ω . A drawing of the coaxial feed is shown in Figure 13-56.

Log-Periodic Folded Slot Antennas

Thorough coverage, including experimentally obtained findings regarding log-periodic folded slot antennas, was in Chapter 14 of the previous edition of this book. Shown in

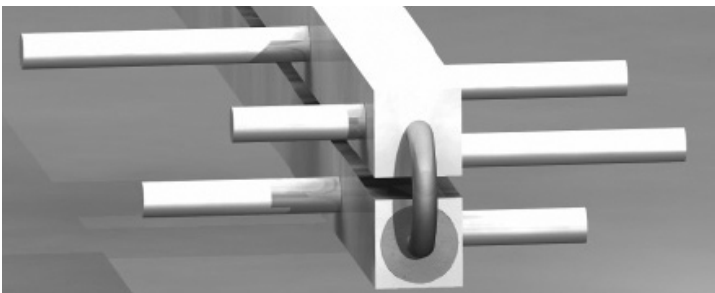


FIGURE 13-56 Coaxial feed arrangement for a log-periodic dipole antenna

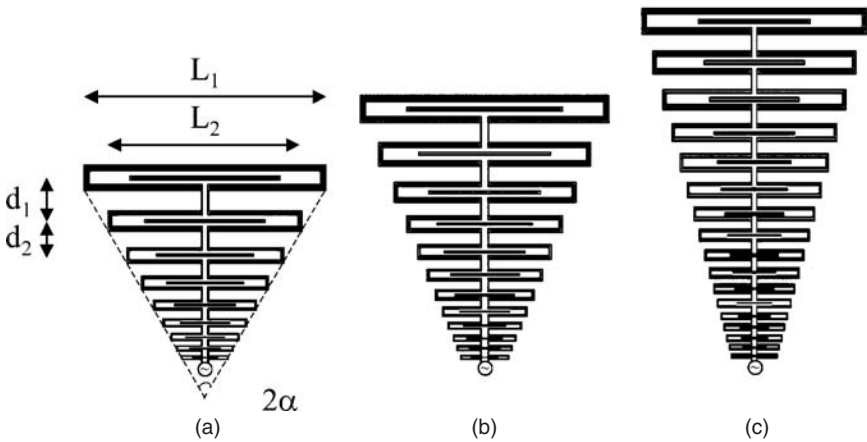


FIGURE 13-57 Log-periodic folded slot antennas with growth rate (τ) of 0.8 (a), 0.85 (b), and 0.9 (c)

Figure 13-57 are several log-periodic folded slot antennas designed for the same frequency range but with three different growth rates and numbers of arms. The shading represents slots etched in a (typically) copper metallization of a thin low-dielectric constant and low-loss substrate. The parameters for these antennas are determined using the steps described by Carrel.⁴⁰ In this case, the designed bandwidth is 4:1, so the number of arms is $N = 9, 12, 17$ for $\tau = 0.8, 0.85, 0.9$, respectively. Between the folded slots are “phasing slots” needed to provide the proper phase delay necessary for end-fire radiation.^{4,49} As with other log-periodic antennas, $\tau = L_{i+1}/L_i$ is the growth rate, and it determines the period of logarithmic repetition in electrical performance. To uniquely specify the antenna geometry, one needs to define the growth rate τ and either σ or α .

As stated earlier, the presence of the phasing slot is critical for end-fire performance of this antenna. If the phasing slot is not used, the growth rate has to be close to 1 in order to reduce gain dropouts and other anomalies in the radiation pattern. It is generally accepted that the radiation patterns are less sensitive for slot lengths when they vary between $\alpha/3$ and 1.5α . Maximum gain for two different lengths of phasing slots is computed using Ansoft Designer and compared with the case without slots (see Figure 13-58).

Three-dimensional radiation patterns for the cases with and without the phasing slot are shown in Figure 13-59. The narrowband gain dropouts, as well as deteriorated front-to-back ratio, multilobes, and so on, are present when the phasing slot is not used. With the phasing slot, the near- and far-field characteristics of the antenna vastly improve. The simulations have shown that the slot lengths around 1α produce the best antenna performance. Measured radiation patterns for a folded slot array are shown in Figure 13-60.

A folded slot log-periodic antenna is most often fed using a coaxial cable, as shown in Figure 13-61. This feed is very similar to an infinite balun, and the parameters of the slot can be easily adjusted for good match. A microstrip feed with a broadband via transition, or a coplanar waveguide line continuing into the log-periodic slot formation, can be easily integrated with this antenna. For conformal integration, the antenna is placed above an absorber-backed cavity.

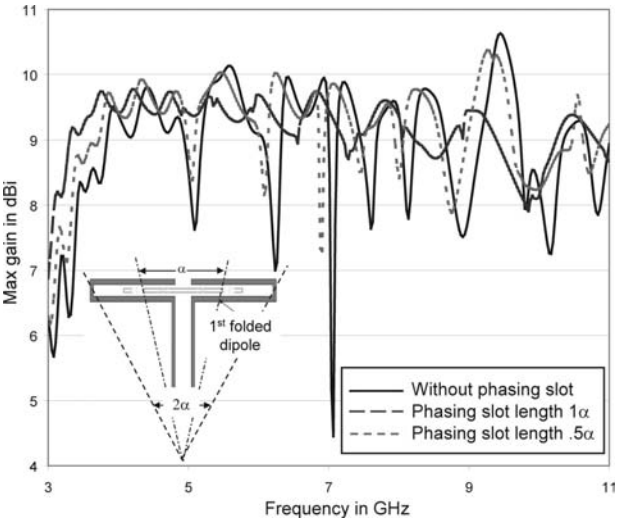


FIGURE 13-58 Maximum gains for different lengths of the phasing slot

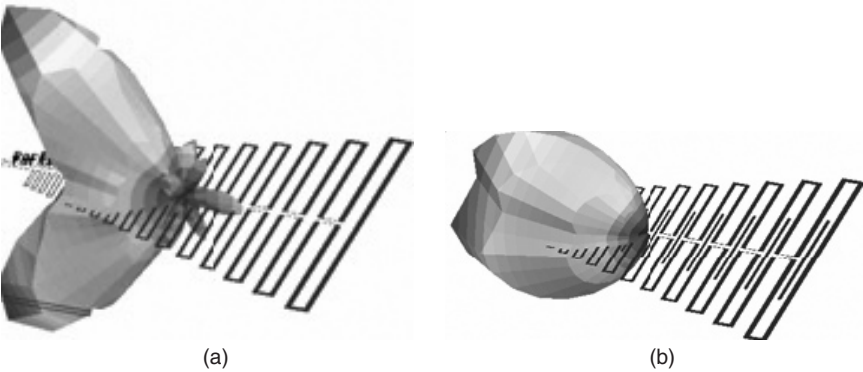


FIGURE 13-59 3D radiation pattern of log-periodic folded slot antennas without (a) and with (b) phasing slots

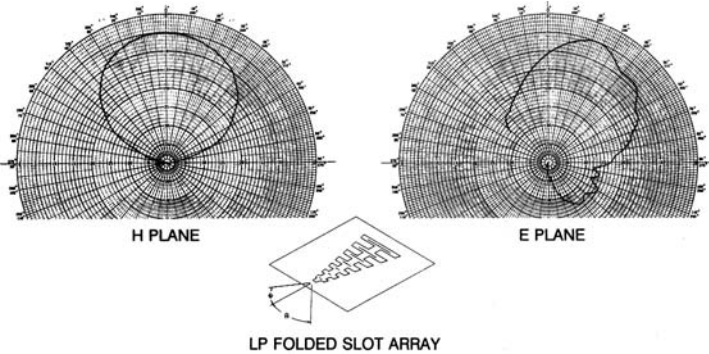


FIGURE 13-60 Measured *E*- and *H*-plane patterns of a folded slot array with phasing slots

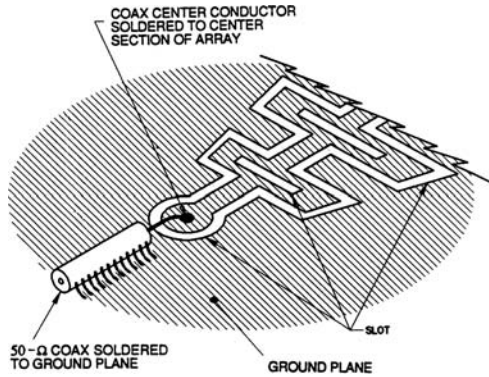


FIGURE 13-61 Log-periodic folded slot array coaxial feed

Log-Periodic Antennas as Reflector Feeds

Log-periodic antennas have been successfully implemented as feeds for reflector antennas.^{44,50,51} The major advantage of these antennas is that they can be designed for a relatively constant impedance and pattern bandwidth, characteristics desired for reduced mismatch and amplitude/phase illumination losses. The major disadvantage is that the phase center moves with frequency—similar to a conical spiral feed. By considering the respective theories of operation, the variation of the phase center with frequency (wavelength) with respect to the tip of (high-frequency end) the conical spiral and non-planar log-periodic can be computed as in Eq. 13-16 and Eq. 13-17, respectively. This variation results in the defocusing losses observed as the reduced gain of the reflector system.⁵²

$$z_{\text{spiral}} = (\lambda/2\pi) \cot \alpha \quad (13-16)$$

$$z_{\text{log-periodic}} = (\lambda/4) \cot \alpha \quad (13-17)$$

Since reflector antennas with the log-periodic feeds are aimed for broadband operation, the positioning of the feed with respect to the focal point of the reflector dish is very important. Before delving into the analysis, note that the phase center variation of the log-periodic from Section 13.3, “Non-Planar Log-Periodic Tooth Antennas,” is somewhat larger than that of the spiral (see Figure 13-62). One can control this spread by choosing an appropriate angle 2α . Sometimes, this parameter varies along the antenna axis. This is done to reduce the antenna height and minimize defocusing losses.

For a specific log-periodic design, the feed position with respect to the reflector’s focal point can be chosen in different ways. In the forthcoming analysis, it was assumed that the defocusing losses are the same at the lowest and the highest frequency of the desired bandwidth. This is the most common approach, and it can be easily shown that the distance between the tip (apex) of the log-periodic feed and the reflector focal point is

$$\Delta = \frac{f_l z_l + f_u z_u}{f_l + f_u} \quad (13-18)$$

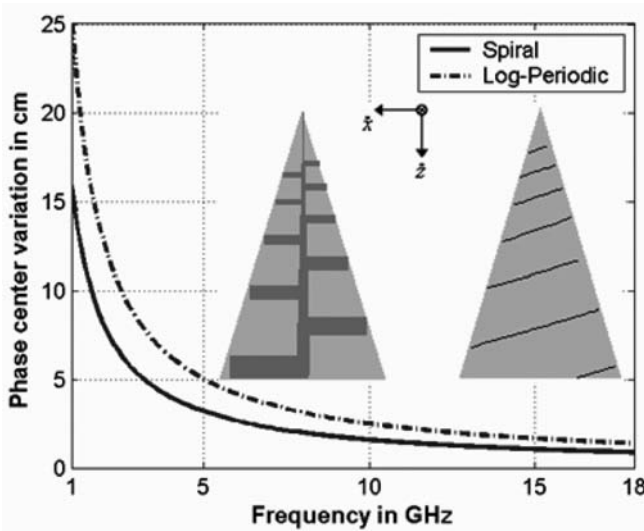


FIGURE 13-62 Comparison of the phase center movement in z -direction versus frequency for conical spiral and non-planar log-periodic feeds designed to operate over the same bandwidth

where z_l and z_f are computed from Eq. 13-17. They represent the locations of the phase center at the lowest and highest frequencies of operation, respectively. Sometimes the position of the feed is chosen such that the defocusing losses are minimized at the highest end of the operating band. In this way, the somewhat larger resistive losses associated with the feeder are compensated for. In applications where high quality performance is required only over a set of narrow (but instantaneous) bands, the location of the feed can be set to minimize the loss at the most critical band.

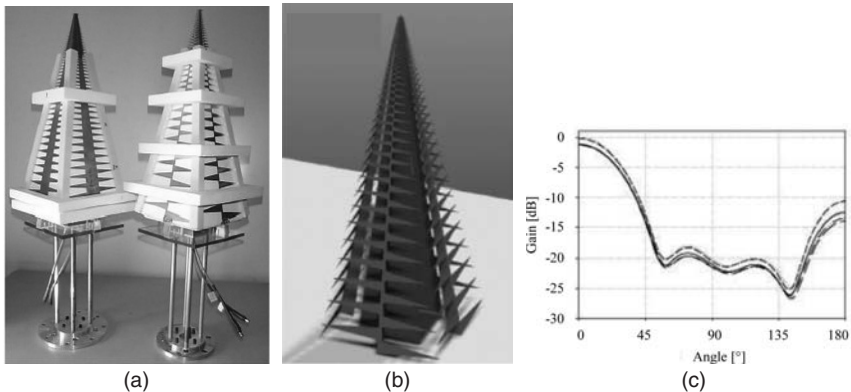
The loss budget of a 10:1 broadband log-periodic tooth-feed for several reflector F/D ratios is shown in Table 13-8. To reduce the losses with an enlarged F/D ratio, the growth rate τ must be increased. The number of elements must also be increased to retain the same bandwidth for optimum illumination. As expected, for lower F/D ratios, the phase error loss (PEL) is increased. Amplitude taper loss (ATL) is also high for low F/D as the feed is too directive and under-illuminates the reflector. The spillover losses (SPL) increase as the F/D increases, and they are the dominant loss mechanism for $F/D > 0.4$. Since the maximum LP directivity is about 10 dB, the reflector becomes over-illuminated, and the ATL is very low. Cross-polarization loss decreases for larger F/D , but primarily due to the more pronounced effects of the feed pattern missing the reflector, thus being counted as spillover loss. Minimum aperture loss is about 2.2 dB, and it is obtained for $0.3 < F/D < 0.4$. For this analysis, the feed was positioned for equal losses at the high and low end of the band. For extremely wide bandwidths, 2α can vary along the feed axis, i.e., it is increased at the lower frequencies to reduce PEL in favor of higher SPL.

High-performance dual-polarized log-periodic feeds with $\psi = 20^\circ$, $2\beta = 3.3^\circ$ and 0.67° , $\tau = 0.975$ and 0.96 , designed for operation over 0.5–11-GHz band are shown in Figure 13-63.⁵⁰ Averaged elevation radiation patterns for five frequencies log-periodically distributed over one period τ are also shown. The patterns are very uniform, with the max gain variation of

TABLE 13-8 Reflector Illumination Losses of Log Periodic Dipole Antennas Designed to Operate over the 10:1 Frequency Range (after Milligan⁵ © J. Wiley & Sons 2005)

Reflector f/D	# Elements	Max. PEL dB	Average ATL dB	Average SPL dB	Cross Pol. Loss dB	Min. Total dB	Max Total dB	Avg. Total dB
0.25	19	0.60	1.52	0.46	0.40	2.32	3.37	2.64
0.30	21	0.44	0.95	0.73	0.36	1.93	2.94	2.20
0.35	23	0.53	0.97	0.71	0.35	1.97	2.95	2.23
0.40	26	0.38	0.44	1.22	0.32	2.05	2.60	2.21
0.45	30	0.40	0.36	1.52	0.28	2.10	2.70	2.30
0.50	33	0.44	0.26	1.63	0.23	2.04	2.80	2.28
0.55	36	0.38	0.20	1.90	0.21	2.19	3.08	2.44
0.60	40	0.38	0.16	2.07	0.17	2.30	3.10	2.53
0.65	43	0.34	0.12	2.33	0.15	2.48	3.30	2.73
0.70	47	0.36	0.11	2.40	0.11	2.50	3.40	2.74
0.75	50	0.36	0.09	2.55	0.09	2.61	3.50	2.85

only 0.75 dB, the first sidelobe level around 18 dB, and a front-to-back ratio of about 13 dB. Measured cross-polarization was around 15 dB within the 13 dB power contour of the beam. By decreasing the boomwidth and growth rate from 3.3° to 0.67° and 0.975 to 0.96, respectively, cross-polarization improved by about 10 dB. The pyramidal shield was added to allow for an undisturbed illumination pattern when the amplifiers are added to the feed point (for reduced feedline losses and their adverse effect on the receiver temperature).

**FIGURE 13-63** Photographs of (a) dual linearly polarized log-periodic feeds, (b) a pyramidal shield, and (c) the roll-angle averaged range profile over one period (after Engargiola⁵⁰ © IEEE 2002)

13.4 DUAL-CIRCULARLY POLARIZED ANTENNAS

Antennas capable of generating an arbitrary polarization over a wide frequency range have been utilized in many applications including electronic warfare, radio astronomy, and remote sensing, just to name a few. Any combination of two orthogonally, linearly polarized radiators with a feed network designed to support different phasing combinations can be utilized for this purpose. A schematic of a feed network utilized to generate left- and right-hand circular polarizations with two sets of orthogonally oriented arm pairs is shown in Figure 13-64. The phase progressions between neighboring arms are 0° , 90° , 180° , 270° (-90°) for RCP and 0° , -90° , -180° , -270° (90°) for LCP. As shown earlier, these two sets of phases contribute to the two ± 1 modes, both having a broadside main beam. Multimode, dual-polarized excitation schemes have also been devised, and they are typically used for monopulse direction finding. For example, a BFN comprised of nine 3-dB magic T's and two 90° hybrids was designed for dual-polarized 2–12-GHz operation in conjunction with an eight-arm monopole log-periodic radiator.⁵³ The stable response with minimum amplitude and phase misbalances over wide bandwidths, as well as the high frequency losses, impose a major challenge in the design of these feeds. Due to their aperture that enables dual-circularly polarized operation (no polarization filtering as in a spiral antenna), the planar embodiments of these antennas are fabricated with an absorber-filled cavity. Any reflective backing, and consequent polarization nondiscriminative coupling to the feeds, severely degrades their impedance and pattern performance. Finally, it is important to note that the cost of the BFN is the major reason for their limited use.

In this section, we discuss several dual-polarized frequency independent configurations with multiple or single feeding arrangements. Specifically, inside-outside fed spirals, modulated armwidth spirals, and sinuous and folded spiral antennas are described. Important structural/geometrical features utilized to realize the dual-polarized performance are outlined. The typical performances and the effects of various parameters have been given. Multiarm dual-polarized and multimode planar log-periodic antennas are discussed in the previous section and will not be addressed here.

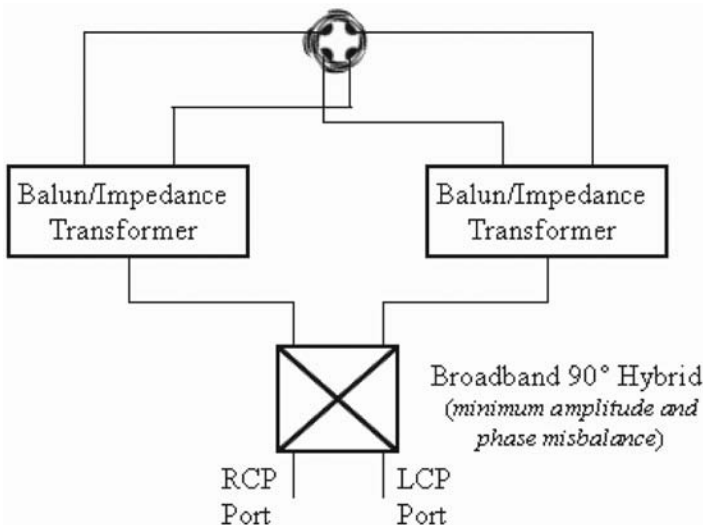


FIGURE 13-64 Circular-polarization feed network for a four-arm, dual-circularly polarized antenna

Inside-Outside Fed Spirals

A spiral can be fed from either the inner ports, outer ports, or both simultaneously. Since the spiral polarization is determined using the right-hand rule, a counterclockwise-wound spiral fed from the inside radiates RCP. The same spiral, when fed from the outer ports, radiates LCP. When access to the spiral is provided from both sides, the antenna becomes a broadband dual-polarized radiator. Feeding from the outside has the additional advantage of allowing ample room to provide impedance matching networks. Stripline or microstrip transitions can be designed to match to 50Ω , and broadband matching for multiple modes can be readily achieved. A major disadvantage of outside feeding is that the highest order mode of excitation will radiate first, thus limiting the effective bandwidth of this type of antenna. To increase the bandwidth, more arms must be added, thus increasing cost due to a more complicated and expensive beamforming network. A higher order, for this structure sometimes also called *parasitic*, mode-free dual-polarized bandwidth of inside-outside fed spirals can be found using expression

$$B \leq \frac{N + 2}{2(M_h + \pi\Delta)} \quad (13-19)$$

where B is the bandwidth ($f_{\text{high}}/f_{\text{low}}$), N is the minimum number of arms, M_h is the highest order mode used, and Δ is the width of the radiation band (typically $\Delta = 0.18$).

Computed bandwidths for multimode operation of inside-outside fed spirals are shown in Table 13-9. To obtain an octave bandwidth dual mode and dual polarization performance at least eight arms are required (1.95:1 bandwidth). For a four-arm spiral, the dual-mode dual-polarization bandwidth is only about 17 percent. By relaxing the requirements for the gain and WoW, the bandwidth will increase.

When the spiral is fed from outside, a continuous conductive ring is required to keep the coaxial feed outside conductors at the same potential. This can be accomplished by using a mounting flange just outside of the spiral cavity. A microstrip or stripline is used to transform the impedance from 50Ω (at the outside port connectors) to match to the spiral arm ending width. This works well for small growth rates toward the spiral perimeter. A drawing of an outer microstrip transition is shown in Figure 13-65.

Measured radiation patterns for an inside-outside fed two-arm loosely wound slot spiral¹⁴ are shown in Figure 13-66. The antenna was fed with a dual-infinite balun feed at both feeding ports. The impedance matching is achieved with the high-impedance large circular element terminating the very narrow, low impedance slotlines constituting the spiral arms. Overlaid are 36 azimuthal cuts taken at frequencies of $1.5f_0$ and $2.5f_0$, where f_0 denotes the

TABLE 13-9 Bandwidth of a Mode-Free Inside-Outside Fed Spiral

N	Mode		
	1	2	3
3	1.6		
4	1.9	1.2	
6	2.5	1.5	1.1
8	3.1	1.9	1.4

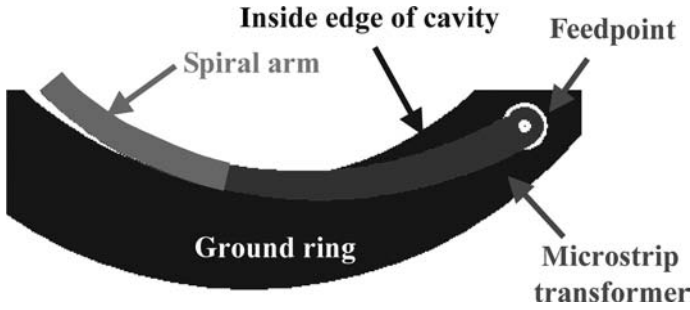


FIGURE 13-65 Outside feed of the spiral using a microstrip transition

frequency associated with the operational wavelength equal to the physical circumference of the antenna. Dual-polarized performance with the broadside beam is clear from the figure. Also, the lower frequency pattern is very symmetric, with very little difference between the two feeds. The higher frequency pattern shows significant degradation in both the symmetry (WoW) and the cross-polarization level. As discussed previously, this is because of the significant attenuation due to radiation through the mode 3 ring. At higher frequencies, the beam squints off center and the mode 3 radiation dominates. A measurement of the far-field phase progression at a constant elevation angle (typically greater than 30°) reveals that for the outside fed spiral, the phase progression of 1080° , associated with mode 3, starts at frequencies close to $3f_0$. For an inside feed, this transition occurs at about $4f_0$ (with this lower efficient slot spiral).

Measured mode content of the radiation pattern of the two-arm inside-outside fed spiral taken at $\theta = 38^\circ$ is shown in Figure 13-67. For the outside feed, the mode 3 radiation starts to dominate the far-field above $3f_0$, while the cross-modal discrimination of greater than 20 dB is seen for frequencies $f > 2f_0$. Note that this slot spiral is not a very efficient radiator when fed inside, thus resulting in even higher mode 3 content than typically measured with well-designed printed spirals.

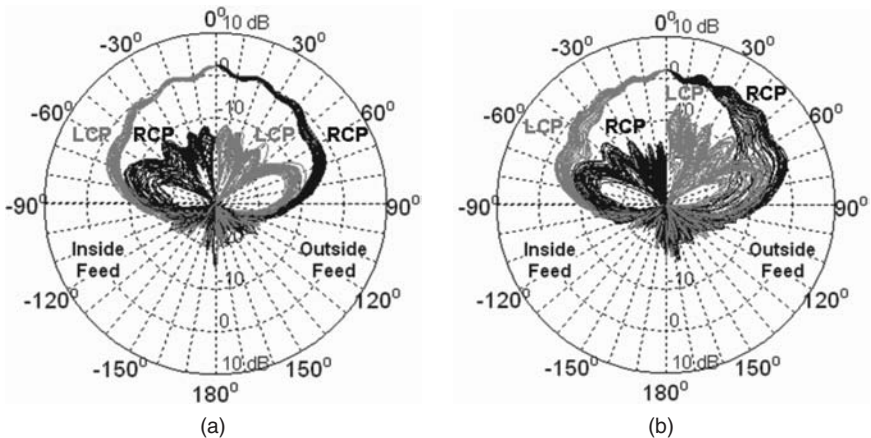


FIGURE 13-66 Measured radiation patterns of a two-arm inside-outside fed spiral taken at $1.5f_0$ (a) and $2.5f_0$ (b). Overlaid are 36 constant ϕ cuts of left- and right-handedness.¹⁴

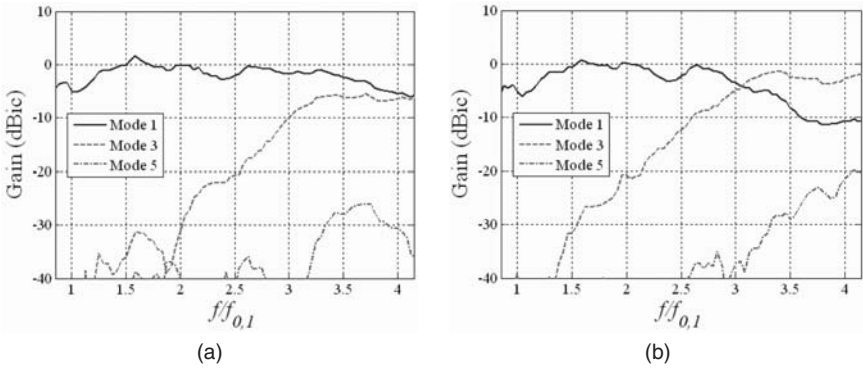


FIGURE 13-67 Measured pattern modal decomposition at an elevation angle of $\theta = 38^\circ$ for the two-arm slot spiral with inside feed (a) and outside feed (b)

Modulated Arm Width Spirals

The MAW spiral¹⁹ was the first center-fed spiral antenna with multi-octave dual-polarized operation. The MAW spiral incorporates arm impedance modulations for creation of bandstop regions. A bandstop, otherwise known as a *band-reflect region*, utilizes a series of half wavelength high/low impedance sections ($1/4$ wavelength each), which provide a broadband method for reflecting the spiral's traveling wave, thus in turn generating the backward oppositely polarized wave. These reflected currents generate modes similar to the outside fed spiral described in the previous section, but without the over-moding limitations. The MAW spiral bandwidth has the same limits as the regular spiral, i.e., the highest frequency is determined by the fineness of the central region, and the lowest limit is determined by the diameter of the spiral. Due to its ability to support both forward and backward waves in dominant active regions, multimode operation can be achieved by increasing the number of arms. For a MAW spiral to operate with m useful modes, the minimum number of arms is $N = 2m + 1$. Thus, for single mode dual-polarized operation, the MAW must have at least three arms. An eight-arm printed/wire MAW is shown in Figure 13-68. Narrow lines represent the high-impedance sections, while the wide strips correspond to the low-impedance sections.

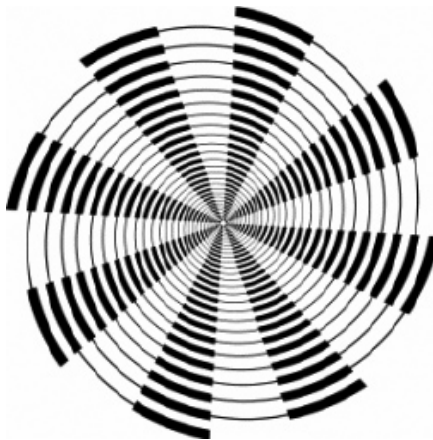


FIGURE 13-68 An eight-arm MAW spiral. The forward wave is RCP; the reflected wave is LCP.

The band-reflect region is set at the circumference of $N/2$ guided wavelengths, where N is the number of arms. A four-arm MAW will reflect the currents at the 2λ circumference, thus limiting this antenna to single mode operation, but preserving the dual circularly polarized capability. To correlate the port phasing with the performance, assume a counterclockwise wound four-arm-printed MAW spiral. This antenna (center-fed) will produce RCP mode 1 radiation for feed phasing vector $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$. Now, if you excite the feed ports with the vector $\{0^\circ, -90^\circ, -180^\circ, -270^\circ\}$, this corresponds to the mode -1 for the RCP spiral. This phase progression is identical to the phase vector $\{0^\circ, 270^\circ, 540^\circ, 810^\circ\}$, which creates a $3 \times 360^\circ$ far-field progression on a $\theta = \text{constant}$ cone and represents the mode 3 phase vector. Thus, the forward traveling wave will pass the mode 1 region without radiating, arrive at its band-stop region at the 2λ circumference, undergo reflection and become a backward (reflected) traveling wave traveling back toward the spiral's center. At the 1λ circumference, phasing is correct to radiate mode -1 and a LCP mode pattern is generated.

The band-reflect section is not required in the low-frequency region. For example, an eight-arm 6-cm diameter MAW will have good mode 2 radiation at 3.3 GHz; however, the mode 4 band reject ring need not exist since the spiral will not radiate mode -2 ($m6$). Since a reflection from the end of the arm is desired, end-loading should not be used for MAW spirals. Also, the inner band reject region need not exist unless it is required for rejecting the highest frequency of the lowest mode. Radiation patterns for forward (RCP) and reflected (LCP) modes 1, 2, and 3 of an eight-arm exponential MAW with EXP = 1.5 taken at 12 GHz are shown in Figure 13-69a. The antenna has 6-cm diameter and the modulation ratio is 11. Thicker lines represent the RCP gain. The main difference is in the deteriorated cross-polarization levels associated with the reflected modes 2 and 3. The difference in mode 1 is insignificant. Overall, the far-field consistency of all modes is acceptable, with small WoW and excellent axial ratio for modes 1 and 2 and a somewhat deteriorated value for mode 3. The input resistance for all modes is given in Figure 13-69b. As with spirals, the modes ± 1 have the highest value of nominal impedance, with very little difference between the forward and reflected modes. Due to the presence of high/low impedance sections, the resistance has highly oscillatory behavior. The nominal mode impedances for forward and reflect modes 1, 2, and 3 are 287Ω , 160Ω , and 127Ω , respectively. This spread of modes and relatively large span of resistance values (from 72Ω to 325Ω) makes beamformer design very challenging. As with spirals, the higher-order modes turn on at the higher frequency.

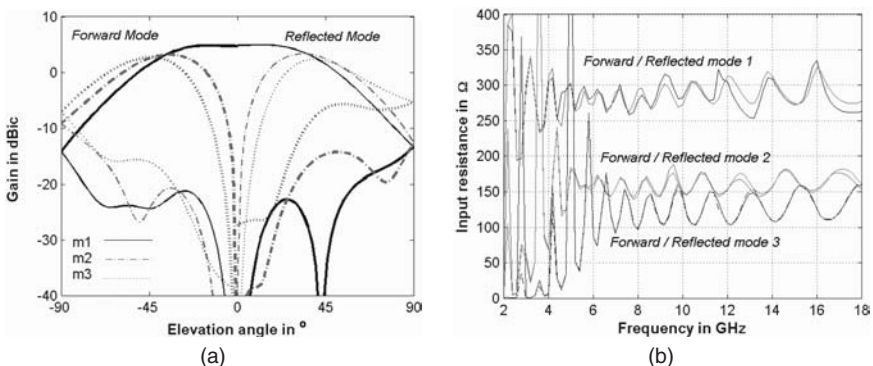


FIGURE 13-69 Radiation patterns for forward and reflected modes of (a) an eight-arm MAW and (b) input resistance vs. frequency. Forward and reflected modes 1, 2, and 3 are considered.

The most important parameter for the performance of a MAW is the *modulation ratio*. It is defined as a ratio between the consecutive wide and thin sections. With an Archimedean MAW, this ratio remains constant throughout the antenna aperture. However, the tightly wound, exponential MAW is typically used due to reduced ringing. Typically, the modulation ratio should be in the 4 to 15 range for good dual-circular polarization of the radiation pattern. A computational study with GNEC is used to demonstrate the effects of this parameter on the performance of a four-arm exponential MAW with $\text{EXP} = 1.5$. The antenna diameter is 7.5 cm, and an ideal beamformer with zero amplitude misbalance and correct phasing is assumed. Modulation ratios in the range from 1 (classical self-complementary equiangular spiral) to 10 are considered. In practice, the fine precision of the feed region and etching finesse are also important to consider. The subset of six different MAWs used in this study is shown in Figure 13-70.

Effect of the modulation ratio on the average cross-polarization level over 2–18GHz and elevation angles $\theta < 30^\circ$ is shown in Figure 13-71a. While the forward mode sees very little variation in the axial ratio, remaining very low throughout the considered range of values, the polarization purity of the reflected mode undergoes significant improvement with increased modulation ratio. Similar conclusions can be drawn for the consistency of the radiation pattern for both modes. Shown in Figure 13-71b are the averaged values for the HPBW over the entire 2–18GHz range, along with their minimum and maximum bounds (associated with the azimuthal variations) plotted against the modulation ratio. While no significant effects are observed with the forward wave, the reflected mode starts to converge when the modulation ratio becomes greater than 8.

Shown in Figure 13-72 are averaged broadside gains from 4–18GHz and WoW at $\theta = 30^\circ$ from 2–18GHz. Forward mode pattern symmetry is very good regardless of modulation ratio. Reflected mode symmetry is somewhat worse, but the average WoW is still about 1 dB over the band. Gain for the forward modes is constant at 5.7 ± 0.3 dBic for all modulation ratios. As expected, reflected mode gain is highly dependent on the modulation ratio. Modulation ratios of at least 8 are required for gains approaching those of the forward modes (within 1 dB).

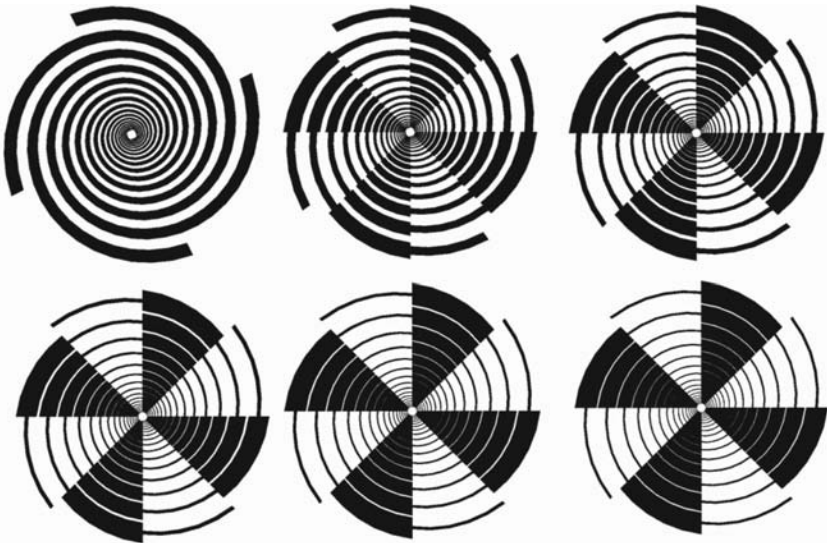


FIGURE 13-70 Four-arm MAWs with modulation ratios (from left to right, top to bottom) of 1, 2, 4, 6, 8, and 10

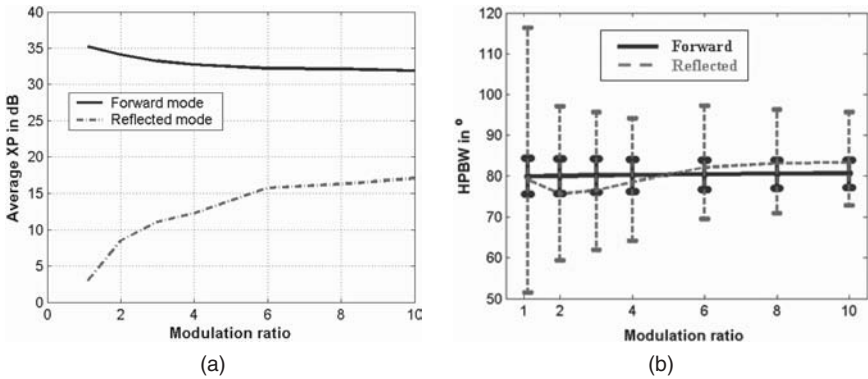


FIGURE 13-71 (a) Average cross-polarization for elevation angles $\theta < 30^\circ$ and (b) 3-dB beamwidth along with the maximum and minimum values for forward and reflected modes in a 2-18 GHz range

Sinuuous Antennas

The traditional sinuous antenna is a broadband, log-periodic structure with $N \geq 4$ arms composed of arcs and bends and whose low- and high-frequency operation limits, when significantly interleaved, are determined by the outer and inner diameters of the antenna, respectively.^{4,41,42} As shown in Figure 13-73, the sinuous configuration is fully depicted by the angles α , or angular span of defining sinuous curve, and δ , or rotation angle needed to outline the sinuous arms, as well as the log-periodic growth rate of τ . If any of these three defining parameters is varied from cell to cell, a quasi-log-periodic sinuous structure is obtained.

Most sinuous antenna research has been devoted to the planar four-arm printed/wire geometries. Since a free-standing antenna radiates bidirectionally, an absorptive cavity or conical projection of the sinuous structure is required for unidirectional operation. Though a lossy cavity backing reduces the antenna efficiency, it is the preferred approach for conformal and flush-mounted applications. Reflective metal backings integrated with the aperture as an empty cavity or cavity with vertical wall-absorber liners, cause strong coupling

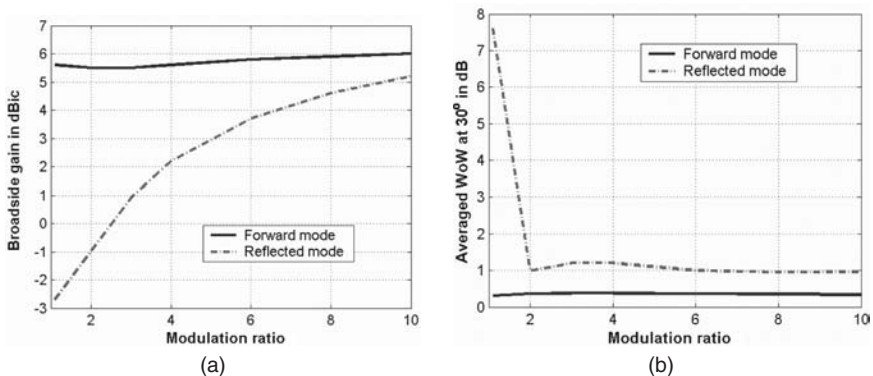


FIGURE 13-72 (a) Averaged broadside gain from 4-18GHz and (b) averaged WoW at $\theta = 30^\circ$ from 2-18 GHz for various modulation ratios and MAW operating in forward and backward modes

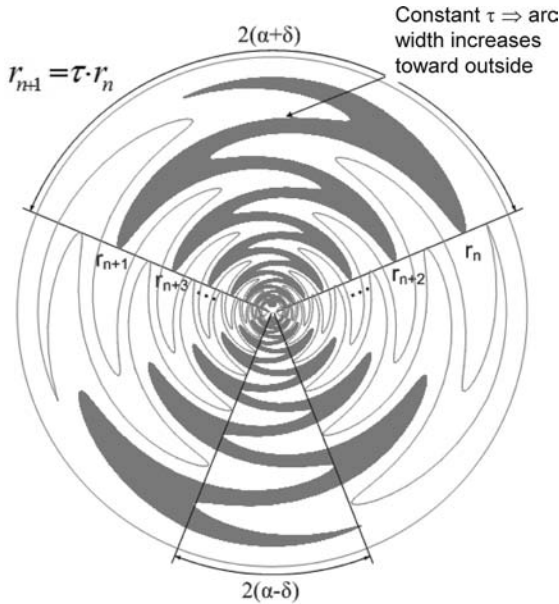


FIGURE 13-73 Sinuous antenna geometry and relevant parameters

into the sinuous feeds and severe mismatch. Additionally, contamination of the radiated far-field occurs as characterized with the gain drop-outs, increased axial ratio, and WoW. Several commercially available sinuous antennas with different mounting arrangements, including lenses, are shown in Figure 13-74.

A tightly coupled sinuous antenna whose arms have significant interlacing—for example, $1.25 \times 2(\alpha + \delta) > 360/N$, where 1.25 denotes the interlacing parameter—has approximately the same frequency limits as a spiral antenna; that is, the sinuous antenna needs to be N wavelengths in circumference to support efficient mode N radiation. In practice, for a given low-frequency limit, the circumference of the sinuous antenna should be increased by a certain value, typically $\lambda/2$, to account for the finite width of the radiation region. Excitation of the sinuous antenna, when operated in a circularly polarized mode, is also identical to the feeding of a spiral antenna. One exception to this similarity is that the sinuous antenna is dual-polarized from a single feed and has



FIGURE 13-74 Sinuous antennas (Courtesy of L-3 Communications)

additional capabilities and, unfortunately, additional restrictions. Due to the alternating clockwise (CW) and counter-clockwise (CCW) wrapping senses in each arm and the inability of the aperture to perform polarization filtering, the sinuous antenna with N arms can support $(N - 1)/2$ broadband modes. Although the overall number of modes on an N -arm sinuous antenna is one half the number of modes on a comparable N -arm spiral, their radiation patterns are very similar. Only mode 1 has a broadside radiating component and all other modes have a null on broadside, with each higher-order mode peaking further from the null. Note that each of the modes on a sinuous antenna can be obtained in both left- and right-handed circular polarization (RCP, LCP). If M modes of operation with the same polarization are desired, an $N = 2M + 1$ arm sinuous antenna is needed. These modes will be excited if equal amplitudes and phase progressions of $-360^\circ nm / N$ are provided at the arm feeds (taken in the CW direction). Here, n is the arm number ($n = 1, 2, \dots, N$), m is the mode number ($m = 1, 2, \dots, \lfloor (N-1)/2 \rfloor$), and N is the total number of arms. If the highest-order mode that can be supported by an arbitrary N -arm sinuous antenna is m ($m > 0$), the following modes can coexist: $-m, -m + 1, \dots, -1, 1, \dots, m - 1, m$, with positive modes denoting the RCP and negative modes denoting the LCP, respectively. Therefore, at least $N = 5$ arms are needed to obtain dual-mode operation, which is important for monopulse direction finding.

To further understand the relation between the number of sinuous arms and the modal spectrum supported by its structure, an arbitrary $N = 4$ arm sinuous geometry will be utilized. A corresponding four-arm CCW wound spiral antenna will support broadband RCP modes M1: $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, M2: $\{0^\circ, 180^\circ, 360^\circ, 540^\circ\} \equiv \{0^\circ, 180^\circ, 0^\circ, 180^\circ\}$, and M3: $\{0^\circ, 270^\circ, 540^\circ, 810^\circ\} \equiv \{0^\circ, -90^\circ, 180^\circ, 90^\circ\}$ with the phase progression (in the CW sense) denoted in brackets $\{\}$. A four-arm sinuous antenna has spiral segments capable of radiating both senses of circular polarization, thus when the mode 2 phase progression is provided at the arm ports $\{0^\circ, -180^\circ, -360^\circ, -540^\circ\} \equiv \{0^\circ, 180^\circ, 0^\circ, 180^\circ\}$, the aperture cannot differentiate between it and the mode -2 phase progression $\{0^\circ, 180^\circ, 360^\circ, 540^\circ\} \equiv \{0^\circ, 180^\circ, 0^\circ, 180^\circ\}$, so the radiating region radiates the mixture of modes 2 and -2 .⁵ The mode 3 phase progression is $\{0^\circ, -270^\circ, -540^\circ, -810^\circ\}$, and the RCP four-arm spiral supports its efficient and broadband radiation (provided the circumference is greater than three wavelengths). If the same phase progression is provided to the sinuous antenna, we notice that it is equal to the progression of $\{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$, which is identical to that of the mode -1 . This mode will be excited first (it radiates from the area the traveling wave encounters before arriving at the mode 3 radiating region) and little energy will be left for mode 3. Thus, as stated previously, for dual-mode operation, a sinuous antenna must have at least five arms, with unique and unambiguous phase progressions for mode 2 $\{0^\circ, -144^\circ, 72^\circ, -72^\circ, 144^\circ\}$, and mode -2 $\{0^\circ, 144^\circ, -72^\circ, 72^\circ, -144^\circ\}$.

Six or eight arms are generally used due to beamformer limitations. An eight-arm sinuous antenna, shown in Figure 13-75, has the additional benefit of supporting a third mode, or modes ± 3 . Typical radiation patterns for the three modes radiated from an eight-arm sinuous antenna are shown in Figure 13-76. Clearly, the previously discussed characteristics of modes 2 and 3 as well as a very good cross-polarization at beam-peak angles are obtained with this structure.

With good quality sinuous designs, currents decay rapidly past the radiation region, and the radiated power of higher-order modes ($m + N, m + 2N, \dots$ and $m - N, m - 2N, \dots$, etc.) is low. Thus, the azimuthal pattern symmetry is usually excellent. As with equiangular spirals, the modal decomposition of the radiated far-field pattern can show the level of modal isolation for each of the sinuous modes. For example, the radiation pattern of mode 1 of an eight-arm sinuous antenna is decomposed into the higher order, contaminating modes. The cross-modal difference with respect to the dominant mode 1 of a first few, most significant modes is shown in Figure 13-77. The cross-modal difference is better than 30 dB for both higher-order modes supported by this antenna (modes -7 and 9),



FIGURE 13-75 An eight-arm self-complementary sinuous antenna

while mode 5 is virtually not excited. Since radiation from the dominant mode region is very efficient, end terminations are rarely needed, as opposed to spirals where some form of lossy arm termination is normally used. Also, for similarly sized and designed feed regions, the high-frequency radiation of a sinuous antenna is usually superior to that of a spiral.

The nominal modal impedance of a free-standing, self-complementary, N -arm printed sinuous antenna can be computed using Deschamps formula Eq. 13-11. Impedance variations from these nominal values can be significant, typically exceeding 2:1 for the mode ± 1 operation. Higher-order modes usually have smaller absolute impedance variations, but these deviations are still greater than their spiral counterparts. If necessary, a spiral can be used in the sinuous interior as a tapered matching section. Typical characteristics of

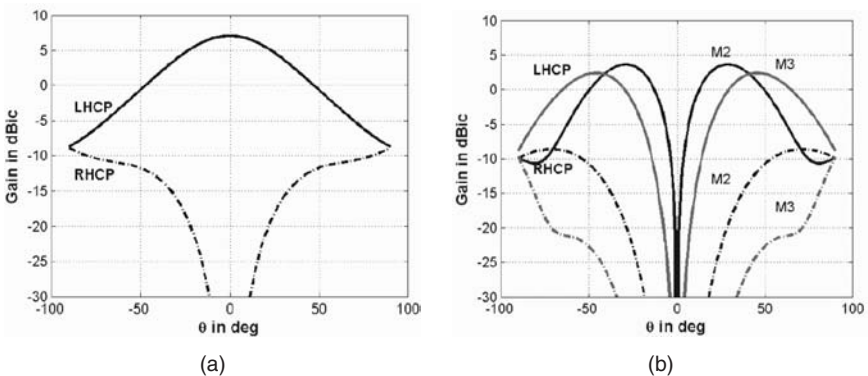


FIGURE 13-76 Radiation patterns for the M1 (a), M2, and M3 (b) of an eight-arm sinuous antenna

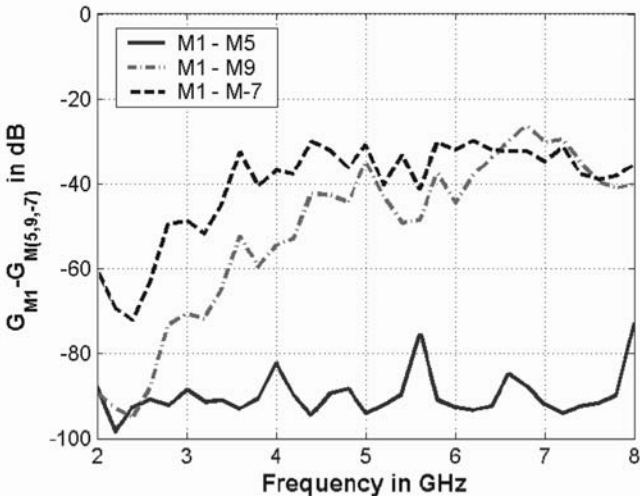


FIGURE 13-77 Higher-order mode gains relative to the gain of M1 for an eight-arm sinuous antenna

self-complementary multimode eight-arm sinuous antennas with different angles 2α are computed using FEKO and the results are shown in Table 13-10. As with other FI antennas, the quality of circular polarization decays with the increased mode number. Also, the antenna gain in general increases with the more pronounced interlacing between the arms. Modes 2 and 3 peak around 34° and 40° , respectively. Note that the lower-frequency end of operation is somewhat extended with the larger arm swings.

Most commercial sinuous antennas are four-arm structures and are designed to operate over different bands. The most common realization is a 5-cm diameter 2–18-GHz antenna operating in a dual-circular mode 1 or dual-linear mode. The dual-circularly polarized sinuous antenna has already been discussed. A linearly polarized version has similar properties to a planar log-periodic antenna. As with good log-periodic design having a typical growth rate of $\tau > 0.8$, there are several arc sections in the active region.

TABLE 13-10 Typical Characteristics at the Midband for an Eight-arm Sinuous Antenna with Various Swing Angles 2α

2α [°]	Mode ± 1			Mode ± 2				Mode ± 3			
	R [Ω]	G [dB]	XP [dB]	R [Ω]	G [dB]	XP [dB]	θ [°]	R [Ω]	G [dB]	XP [dB]	θ [°]
22.5	249	5.0	51.8	143	3.6	25.1	34	111	3.4	18.4	38
37.5	244	5.2	52.0	140	3.6	26.2	36	110	3.4	13.9	40
52.5	252	5.6	49.7	149	3.7	25.8	34	115	3.2	18.9	40
67.5	256	5.7	50.5	150	3.9	25.5	32	120	3.5	17.2	38
82.5	255	5.9	49.5	151	4.2	30.1	36	120	3.5	19.7	40

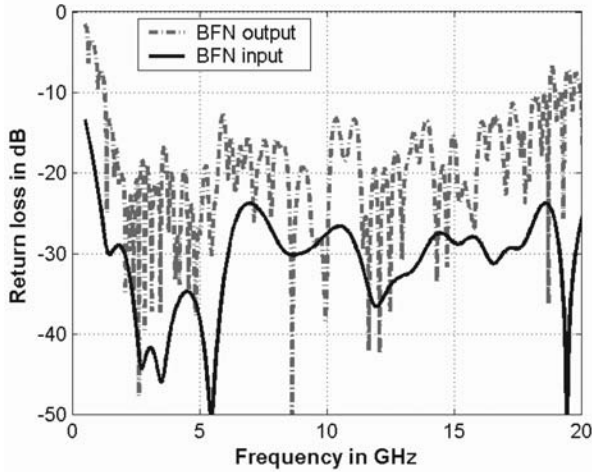


FIGURE 13-78 Measured return loss of a well-designed 5-cm cavity-backed sinuous antenna at the input and output of the beamformer

Opposite arms are fed antiphase ($0, 180^\circ$) and the two orthogonal arm-pairs of a four-arm sinuous antenna form two linearly polarized radiators. A small shift in the polarization vector with changing frequency is present as a consequence of the oscillating sinuous form of the structure. The polarization swings back and forth about the arms' central axis. When operating in circular polarization (with the addition of a 90° hybrid), this polarization oscillation or wobble does not degrade the polarization purity since each arm has identical rotation.

The return loss of a well-designed four-arm sinuous antenna is shown in Figure 13-78. Highly oscillatory behavior is due to the reflections inherent to the BFN, and after gating those out, a very smooth over 20-dB return loss is measured. The coupling between the two channels can be as high as -10 dB, while the losses in the BFN are typically below 1.5 dB at the highest end.

The performance of commercially available sinuous antennas, shown in Figures 13-79 and 13-80, clearly demonstrates high-quality wideband operation of these antennas.

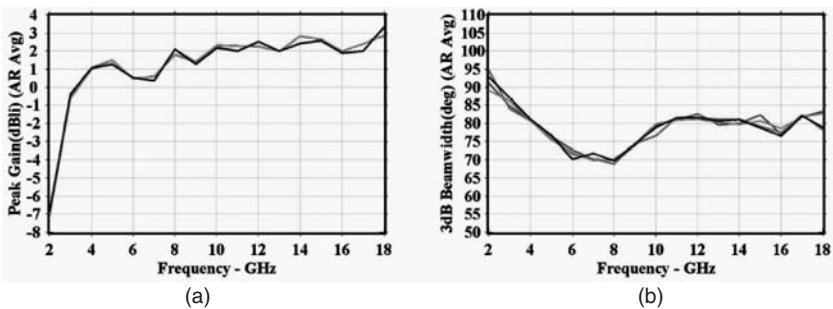


FIGURE 13-79 Measured peak gain (a) and 3-dB beamwidth (b) for two channels in two principal planes of a 5.5-cm sinuous antenna (Courtesy of L-3 Communications)

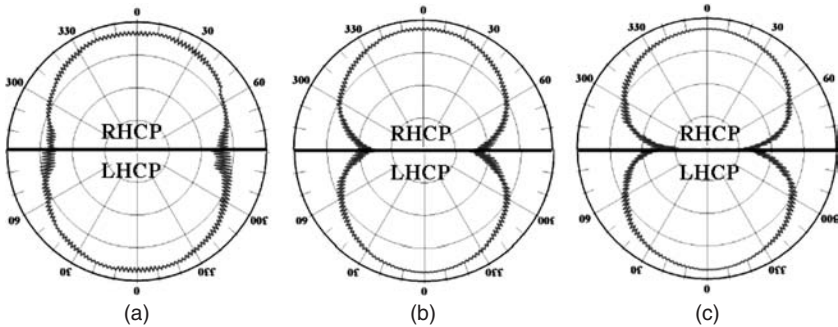


FIGURE 13-80 Measured, rotated linear-polarization radiation patterns for two channels of a 5.5-cm sinuous antenna at (a) 2 GHz, (b) 9 GHz, and (c) 18 GHz (*Courtesy of L-3 Communications*)

Folded Spirals

A special case of sinuous antenna is the folded spiral, constructed as a spiral that is “folded” back on itself. The angular arm expansion of the folded spiral is the same as that of a sinuous antenna, and the main difference is in the shape or curvature of the arm bends. More specifically, the bends of a sinuous structure are rounded as opposed to the sharp tips of a folded spiral. The performance of the two antennas is similar, and often the folded spirals are referred to as sinuous antennas. Thus any reference to one can be equally applied to the other antenna.

Conical Structures

The unidirectional radiation from dual-polarized antennas follows the same rules and guidelines as shown for the conical spiral antenna (Section 13.2). Decreasing the cone angle results in a longer active region and will increase the forward radiation while suppressing the backward radiation (higher F/B ratio). Antenna parameters such as number of arms, radiation regions, modes, and so on, work the same as discussed in previous sections. The planar antenna geometry (spiral, sinuous, MAW, and so on) is projected on to a cone to form the conical antenna. As with conical spirals, the antenna phase center will move along the z -axis with frequency. Radiation will occur where the circumference is N guided wavelengths or the individual elements are resonant length. Sinuous and LP tooth geometries have similar radiation proportions for both polarizations (RCP and LCP). The MAW and inside-outside fed spirals have reduced directivities in the reflected (or outside fed) polarizations but might still be useful for certain applications.

Acknowledgments

The authors would like to thank Mr. Michael Buck and Mr. Neill Kefauver from the University of Colorado at Boulder, Mr. Nathan Stutzke from Ball Aerospace, and Mr. David DelRio from the University of Puerto Rico for their contributions to this chapter.

REFERENCES

1. V. H. Rumsey, "Frequency Independent Antennas," *IRE Int. Conv. Rec.*, vol. 5 (March 1957): 113–118.
2. V. H. Rumsey, *Frequency-Independent Antennas* (New York and London: Academic Press, 1966).
3. P. E. Mayes, "Frequency-Independent Antennas: Birth and Growth of an Idea," *IEEE Antennas Propagat. Soc. News.* (August 1982): 4–8.
4. R. H. DuHamel and J. P. Scherer, "Frequency Independent Antennas," Chap. 14, *Antenna Engineering Handbook, Third Edition* (New York: McGraw-Hill, 1993).
5. T. A. Milligan, *Modern Antenna Design* (New York: J. Wiley & Sons, 2005).
6. C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd Ed. (New York: John Wiley & Sons, Inc., 1996).
7. D. S. Elliott, "A View of Frequency Independent Antennas," *Microw. Journ.* (December 1962): 61–68.
8. R. H. DuHamel and D. E. Isbell, "Broadband Logarithmically Periodic Antenna Structures," *IRE Int. Conv. Rec.*, vol. 5 (March 1957): 119–128.
9. E. M. Turner, Spiral Slot Antenna, (1958): U.S. Pat. 2,863,145.
10. J. D. Dyson, "The Equiangular Spiral Antenna," *IRE Trans. Ant. Propagat.*, vol. AP-7 (April 1959): 181–187.
11. J. D. Dyson, "The Unidirectional Equiangular Spiral Antenna," *IRE Trans. Ant. Propagat.*, vol. AP-7 (October 1959): 329–334; "The Characteristics and Design of the Conical Log-Spiral Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-13 (July 1965): 488–499.
12. J. Kaiser, "The Archimedean Two-Wire Spiral Antenna," *IRE Trans. Antennas Propagat.*, vol. AP-8 (May 1960): 312–323.
13. R. Corzine and J. Mosko, *Four-Arm Spirals* (Norwood, MA: Artech House, 1990).
14. D. S. Filipovic, N. A. Stutzke, M. C. Buck, Q. Mu, and N. W. Kefauver, "Unconventional Spiral Antennas and Arrays," *Proc. 29th Ant. Applic. Sympos.* (September 2005): 378–407.
15. J. Wang and V. Tripp, "Design of Multi-Octave Spiral-Mode Microstrip Antennas," *IEEE Trans. Ant. Propagat.*, vol. 39 (March 1991): 332–335.
16. H. Nakano, Y. Shinma, and J. Yamauchi, "A Monofilar Spiral Antenna and Its Array above a Ground Plane-Formation of a Circularly Polarized Fan Beam," *IEEE Trans. Ant. Propagat.*, vol. 45 (October 1997): 1506–1511.
17. P. E. Mayes and J. D. Dyson, "A Note on the Difference Between Equiangular and Archimedes Spiral Antennas," *IRE Trans. Microw. Theory and Techn.*, vol. 9 (March 1961): 203–204.
18. J. Donnellan, "Second-Mode Operation of the Spiral Antenna," *IRE Trans. Antennas Propagat.*, vol. AP-8 (November 1960): 637.
19. P. G. Ingerson, "Modulated Arm Width (MAW) Log-Spiral Antennas," *20th Annu. Symp., USAF Ant. Res. Dev. Prog.*, October 1970.
20. B. Cheo, V. Rumsey, and W. Welch, "A Solution to the Frequency-Independent Antenna Problem," *IRE Trans. Ant. Propagat.*, vol. AP-9 (November 1961): 527–534.
21. R. Sivan-Sussman, "Various Modes of the Equiangular Spiral Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-11 (September 1963): 533–539.
22. C-C Chen and J. L. Volakis, "Spiral Antennas: Overview, Properties, and Miniaturization Techniques," *Printed Antennas for Wireless Communications*, Chap. 3 (New York: Wiley InterScience, April 2007).
23. J. L. Volakis, M. W. Nurnberger, and D. S. Filipovic, "A Broadband Cavity-Backed Slot Spiral Antenna," *IEEE Antennas Propagat. Mag.*, vol. 43 (December 2001): 15–26.
24. T. Milligan, "Parameters of Multiple-Arm Spiral Antenna from Single-Arm Measurements," *IEEE Ant. Propagat. Mag.* vol. 40 (December 1998): 65–69.

25. G. Deschamps, "Impedance Properties of Complementary Multiterminal Planar Structures," *IRE Trans. Ant. Propagat.*, vol. AP-7 (December 1959): 371–378.
26. J. Huffman and T. Cencich, "Modal Impedances of Planar, Non-complementary, N-fold Symmetric Antenna Structures," *IEEE Ant. Propagat. Mag.*, vol. 47 (February 2005): 100–116.
27. T. P. Cencich, J.A. Huffman, and D. Walcher, Simultaneous Mode Matching Feedline, (2003): U.S. Pat. 6,549,175.
28. F. O'Hara, J. Toth, J. Weiss, and A. Chinchillo, Matching Section for Multi-arm Spiral Antenna, (August 2, 1983): U.S. Pat. 4,396,921.
29. D.F. Bowman, "Impedance Matching and Broadbanding," *Antenna Engineering Handbook, Second Edition*, Chap. 43 (New York: McGraw-Hill, 1984).
30. N. Marchand, "Transmission Lines Conversion Transformer," *Electronics*, vol. 17 (1944): 142–145.
31. N. Stutzke and D. S. Filipovic, "Four-Arm 2nd Mode Slot Spiral Antenna with Simple Single-Port Feed," *IEEE Ant. Wirel. Propagat. Let.*, vol. 4 (2005): 213–216.
32. M. Buck and D. Filipovic, "Spiral Cavity Backing Effects on Pattern Symmetry and Modal Contamination," *IEEE Ant. Wirel. Propagat. Let.*, vol. 5 (2006): 243–246.
33. M. W. Nurnberger and J. L. Volakis, "New Termination for Ultrawide-band Slot Spirals," *IEEE Trans. Antennas Propagat.*, vol. 50 (January 2002): 82–85.
34. D. S. Filipovic and J. L. Volakis, "A Broadband Meanderline Slot Spiral Antenna," *IEE Proc. Microw. Antennas Propagat.*, vol. 149 (December 2002): 98–105.
35. W. Curtis, "Spiral Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-8 (May 1960): 298–306.
36. T. Ozdemir, J. L. Volakis, and M. W. Nurnberber, "Analysis of Thin Multioctave Cavity-Backed Slot Spiral Antennas," *IEE Proc. -Microw. Antennas Propagat.*, vol. 146 (December 1999): 447–454.
37. C. Fumeaux, D. Baumann, and R. Vahldieck, "Finite-Volume Time-Domain Analysis of a Cavity-Backed Archimedean Spiral Antenna," *IEEE Trans. Antennas Propagat.*, vol. 54 (March 2006): 844–851.
38. H. Nakano, K. Nogami, H. Mimaki, and J. Yamauchi, "A Spiral Antenna Backed by a Conducting Plane Reflector," *IEEE Trans. Antennas Propagat.*, vol. AP-34 (June 1986): 791–796.
39. R. H. DuHamel and F. R. Ore, "Logarithmically Periodic Antenna Designs," *IRE Int. Conv. Rec.*, vol. 6 (March 1958): 139–151.
40. R. L. Carrel, "The Design of Log-Periodic Dipole Antennas," *IRE Int. Conv. Rec.*, vol. 9 (March 1961): 61–75.
41. R. H. DuHamel, Dual Polarized Sinuous Antennas, (1987): U.S. Pat. 4,658,262.
42. T. T. Chu and H. G. Oltman, "The Sinuous Antenna," *MSN & Comm./Tech.*, vol. 18 (June 1988): 1–4.
43. T. Hertel and G. Smith, "Analysis of Conical LP Antennas," *Microw. Optic. Tech. Let.*, vol. 36 (January 2003): 28–32.
44. R. H. DuHamel and F. R. Ore, "LP Feeds for Lens and Reflectors," *IRE Int. Conv. Rec.*, vol. 7 (March 1959): 128–137.
45. D. E. Isbell, "Log-Periodic Dipole Arrays," *IRE Trans. Antennas Propagat.*, vol. AP-8 (May 1960): 260–267.
46. P. C. Butson and G. T. Thompson, "A Note on the Calculation of the Gain of Log-Periodic Dipole Antennas," *IEEE Trans. Ant. Propagat.*, vol. AP-24 (January 1976): 105–106.
47. G. De Vitto and G. B. Stracca, "Comments on the Design of Log-Periodic Dipole Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (May 1973): 303–308.
48. G. De Vitto and G. B. Stracca, "Further Comments on the Design of Log-Periodic Dipole Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (September 1974): 713–718.
49. J. W. Greiser, "A New Class of Log-Periodic Antennas," *IEEE Proc.*, vol. 52 (May 1964): 617.
50. G. Engargiola, "Non-Planar Log-Periodic Antenna Feed for Integration with a Cryogenic Microwave Amplifier," *Proc. IEEE Ant. Propagat. Symp.*, vol. 4 (June 2002): 140–143.

51. R. Olsson, P-S. Kildal, and S. Weinreb, "The Eleven Antenna: a Compact Low-Profile Decade Bandwidth Dual Polarized Feed for Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. 54 (February 2006): 368–375.
52. R. Dybdal, "Defocusing Loss for a Log-Periodic Fed Reflector," *IEEE Trans. Antennas Propagat.*, vol. 33 (July 1985): 809–812.
53. R. H. DuHamel, M. E. Armstrong, and W. F. Pedler, "Dual-Polarization Monopulse Antennas Utilizing Conical Arrays of Log-Periodic Antennas," *Proc. IEEE Ant. Prop. Symp.*, vol. 4 (December 1966): 419–424.

Chapter 14

Horn Antennas

Trevor S. Bird

CSIRO Australia

Allan W. Love

Consultant

CONTENTS

14.1	INTRODUCTION	14-2
14.2	RECTANGULAR HORNS	14-9
14.3	CIRCULAR HORNS	14-28
14.4	CORRUGATED HORNS	14-42
14.5	MISCELLANEOUS HORN STRUCTURES	14-60

14.1 INTRODUCTION

Types of Horns and Applications

Electromagnetic horn antennas make a transition from waves propagating in a transmission line, usually a waveguide, and the launching of waves into another medium such as free-space or underground. They occur in a variety of shapes and sizes to fulfill many practical applications. These vary from communication systems, electromagnetic sensing, radio frequency heating, nondestructive testing and evaluation, biomedicine, and as a reference source for other antenna testing. Horns are used as feeds for other antennas such as reflectors, lenses, and compound antennas. Their widespread use stems from their simple, solid geometry and excellent performance when beam directivity is called upon. A summary of the many types of horns and typical applications with some references is given in Table 14-1. It is clear that one horn does not fit all situations—although their main purpose is to be a source of directive power or a directive receiver. For example, in communications and radio astronomy, a horn should be very low loss and be immune to extraneous signals including noise. Also, in some situations, it may be necessary that the horn operate over an octave or more of frequencies, such as in ultrawideband or pulse signal applications. Further, horns may be required to handle two orthogonal polarizations and minimize interference between them. These and other properties will be highlighted in this chapter. A collection of important historical papers dealing with horn antennas can be found in *Electromagnetic Horn Antennas*.¹

Some examples of these horns are pictured in Figure 14-1. Probably the most commonly occurring horn is the rectangular pyramidal family. The pyramidal horn shown

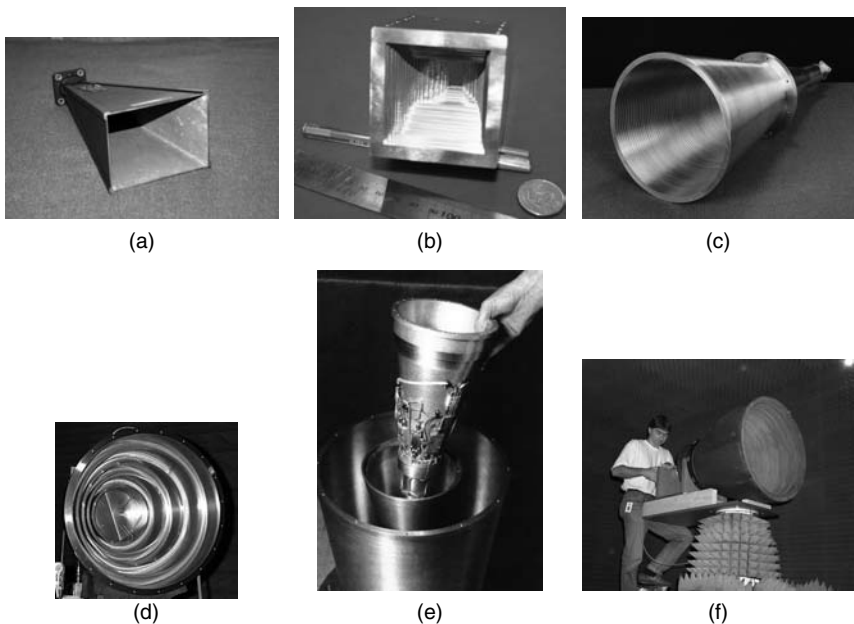


FIGURE 14-1 A selection of different types of horn antennas: (a) pyramidal horn; (b) profiled rectangular horn; (c) circular horn with axial corrugations and dual-probes for circular polarization; (d) conical corrugated horn; (e) dual-band coaxial horn with tracking; (f) compact profiled corrugated horn (Courtesy © CSIRO Australia 2007)

TABLE 14-1 Types of Horn Antennas and Their Applications

Classification	Application
<i>Rectangular</i>	
Smooth wall ¹⁻⁵	Standard gain, testing, sensing
Corrugated ⁶⁻¹⁰	Radar, heating
Dielectric-lined ¹¹⁻¹²	Radar, satellite
Profiled ^{1,13-15}	Satellite
Multimode ¹⁶⁻²⁰	Radar, satellite, radio astronomy, tracking
Self supporting ²¹⁻²³	Reflector feed
Ridged—double, quad ^{1,24-27}	Ultrawideband, radar, sensing
<i>Circular</i>	
Smooth-wall conical ^{1-2,28-33,110}	Feeds, testing, sensing
Corrugated ^{1,34-45}	Satellite, space, radio astronomy, tracking
Scalar ^{1,34,39-40,45}	Satellite, space, radio astronomy
Dielectric-lined ⁴⁵⁻⁵¹	Wideband, radio astronomy
Coaxial ^{55,52-58}	Multiband feed, probe, nondestructive testing, medical
Self supporting ⁵⁹⁻⁶⁴	Reflector feeds
Profiled ⁶⁵⁻⁷⁰	Satellite, radio astronomy
Special flange ^{39,71-73}	Feeds, medical
Longitudinal corrugation ⁷⁴⁻⁷⁶	Satellite, radio astronomy
Multimode ^{1,45,77-85}	Radar, satellite, radio astronomy
Ridged—double, quad ⁸⁶⁻⁸⁸	Ultrawideband, radar, sensing
<i>Other Types</i>	
Diagonal ^{1,89-90}	Radio astronomy, sensing
Elliptical—all types ⁹¹⁻⁹⁹	Reflector feed, radar, satellite
Superelliptical ¹⁰⁰	Feeds
Oblique aperture ¹⁰¹	Feeds
Dielectric—all types ¹⁰²⁻¹⁰⁸	Direct radiator

in Figure 14-1a is used as a primary gain standard since its gain may be calculated to within 0.1 dB by knowing its dimensions. Independent control of the beamwidths in the two principal planes is possible by varying the rectangular-aperture dimensions. A special case is sectoral horns, which are flared in only one plane. They radiate fan-shaped beams, broad in the plane orthogonal to the flare. Other rectangular horns (Figure 14-1b) can have a nonlinear profile to increase efficiency and lower sidelobes compared with a linear flare.

Circular geometries (for example, Figure 14-1c and d) are also in widespread use, with other shapes such as elliptical and superelliptical¹⁰⁰ providing additional variation when

required for beam-shaping. The conical horn is, by virtue of its axial symmetry, capable of handling any polarization of the exciting dominant TE_{11} mode. It is particularly well-suited for circular polarization. Despite its axial symmetry, however, the beamwidths in the principal planes are generally unequal. Like the pyramidal horn, the conical horn may be used as a primary standard since its gain is accurately calculable. Also, profiling will improve performance.

The dual-mode and corrugated (hybrid-mode) conical horns shown in Figure 14-1c, d, and f overcome the lack of axial symmetry in the radiation pattern, and at the same time, achieve diminution of cross-polarized radiation in the intercardinal planes to a very low level. In both cases, the aperture field of the horn is modified in such a way as to produce radiation patterns with axial symmetry. A horn with axial corrugations, shown in Figure 14-1c, converts part of the incident TE_{11} mode in circular waveguide into TE_{11} mode in short-circuited coaxial waveguide to generate sufficient TM_{11} mode in the horn to produce patterns with good axial symmetry. The coaxial horn in Figure 14-1e allows two frequency bands to be used simultaneously and also provides tracking in the highest band.

Horns can have metallic walls, be composed of dielectric material, or employ combinations. In this chapter, we concentrate on horns with metallic sidewalls, although in the last section we review briefly dielectric-only horns (see Table 14-1).

Important Characteristics of Horns

Performance determines the selection of a horn for a specific application. This selection is based on a number of quantities and qualities, usually

- Bandwidth properties
- Radiation pattern
- Gain and efficiency
- Phase center
- Polarization
- Input match
- Fabrication and cost

These attributes will be briefly considered in turn.

Bandwidth Properties The chosen application normally includes a range of frequencies. Horn antennas are often differentiated by their ability to cover a narrow or wide frequency range. This is sometimes expressed as a percentage bandwidth, which is

$$\% \text{ bandwidth} = 200 \frac{(f_{\max} - f_{\min})}{(f_{\max} + f_{\min})} \quad (14-1)$$

where f_{\min} and f_{\max} are the minimum and maximum operating frequencies. Another way to express the frequency properties is as a ratio such as f_{\max}/f_{\min} :1. For example, a corrugated horn could be designed to operate from 10 to 21 GHz. That is, the percentage bandwidth is about 71% or the frequency ratio is 2.1:1.

Radiation Pattern The radiated fields from a general antenna lie on the surface of an imaginary sphere enclosing it. In many applications, it is usual to assume the radius of this sphere is large and a far-field approximation can be employed (see Chapter 1).

This means in a spherical coordinate system, the radial components are negligibly small, and the remaining field components are in the θ - and ϕ -directions. In the far-field the electric \mathbf{E} and magnetic \mathbf{H} fields are related by the wave impedance of free-space η (≈ 376.7 ohms) as follows: $\mathbf{H} = (\hat{r} \times \mathbf{E})/\eta$. The radiated power density in the outward radial direction is, therefore,² $P_r = |\mathbf{E}|^2/2\eta$. The radiation pattern is the distribution of this power density over the far-field sphere. The general properties of antenna radiation patterns are described elsewhere. Particular to horns is that the power density is directed in front of the horn, often with specifications on the sidelobe and backlobe levels. Sections of the radiation pattern for horns are usually given as a function of the elevation variable θ at a constant azimuth angle ϕ . For linear polarization, the \mathbf{E} -, \mathbf{H} - and intercardinal (± 45 -degree) planes are the most useful constant ϕ -pattern cuts through the two-dimensional pattern. The main beam (usually the axial direction) can be described by the half-power beamwidth (HPBW) and the angle between a lower power level relative to the peak, say -10 or -12 dB down.

Gain and Efficiency Antenna gain is defined in another chapter of this handbook. Several different types of gain are used to describe horn antennas. The most commonly accepted definition of gain is provided by the IEEE (Std. 145 – 1993 §2.165), which is based on the power accepted by the antenna. This definition removes the power reflected from the input from the definition. Another definition, called *absolute gain*, is also used.² It retains the reflected power from the input, as this is an intrinsic property of the antenna. For horn antennas, there is often very little difference between these two definitions because the reflection coefficient is small. The absolute gain, G , differs from the IEEE definition by the factor $1/(1 - |\Gamma_{in}|^2)$, where Γ_{in} is the input reflection coefficient. Absolute gain will be used throughout this chapter.

For horn antennas, and aperture antennas in general, the aperture efficiency is another useful measure for comparing antennas. Aperture efficiency in the direction (θ_m, ϕ_m) is the ratio of the gain of the antenna in that direction and the gain of an ideal uniformly illuminated aperture given by

$$\eta_a = \frac{\lambda^2}{4\pi A} G(\theta_m, \phi_m) \quad (14-2)$$

where A is the total area of the aperture, λ is the wavelength at the operating frequency, and G is the gain function (Chapter 1). Many applications require the aperture efficiency on axis, i.e., $(\theta_m, \phi_m) = (0, 0)$.

Phase Center This is the apparent center of the spherical waves that emanate from the horn at a given radial distance (usually the far-field, although for feeds this will be the average distance to the reflector). At the phase centre and in the far-field region, the phase response of the radiation pattern in the vicinity of the main beam will be reasonably constant. The phase center can be calculated from the radiation pattern¹⁰⁹ or measured by adjusting the center of rotation until the measured phase of the pattern is constant across the main beam, say between the -12 dB points—or a wider range when the horn is used as a feed. The phase center can vary significantly with frequency when the pattern angles extend across the main beam. The phase center d_{pc} of a particular radiation pattern cut $\phi = \phi_i$, having a phase function $\Phi(\theta, \phi_i)$, can be calculated from

$$kd_{pc}(\phi_i) = \frac{C_{10}C_{01} - C_{11}C_{00}}{C_{01}^2 - C_{02}C_{00}} \quad (14-3)$$

where

$$C_{mn}(\phi_i) = \int_{\theta_1}^{\theta_2} \Phi(\theta, \phi_i)^m \cos^n \theta d\theta$$

$k = 2\pi/\lambda$ and $\theta = [\theta_1, \theta_2]$ is the angular extent of the pattern. Equation 14-3 was derived by a least-squares comparison of the horn phase function and the ideal phase over the specified angular range. Note that d_{pc} is given relative to the origin of the spherical coordinate system. It is easily seen that when the phase function has positive curvature relative to the origin, the phase center is in front of the horn, and when the curvature is negative, the phase center is inside the horn. The overall phase center of the horn can be estimated by averaging the phase centers of several pattern cuts.

Polarization The antenna radiation on the far-field sphere can be resolved into two components. The largest component in the direction of the desired polarization is called copolarization, and the orthogonal component is cross-polarization. Polarization performance measures of particular relevance to horns are peak cross-polarization for linear polarization and axial ratio for circular polarization (see Chapter 1). Measures of low cross-polarization are low peak levels in the intercardinal planes and even lower levels in the principal planes. There are several ways of measuring cross-polarization. For horns, the copolar pattern is obtained by aligning the test antenna on boresight with the polarization of the distant source antenna. The test antenna is then rotated about the phase center while maintaining this alignment. The signal that is measured is the copolar pattern. The cross-polar pattern is obtained by rotating the transmitter through 90° and repeating the radiation pattern measurement. The received signal is the cross-polar pattern. In this definition, the co- and cross-polarized field components, E_{cp} and E_{xp} , may be obtained from the spherical field components using

$$\begin{bmatrix} E_{cp} \\ E_{xp} \end{bmatrix} = \begin{bmatrix} \cos(\phi - \phi_o) & -\sin(\phi - \phi_o) \\ -\sin(\phi - \phi_o) & -\cos(\phi - \phi_o) \end{bmatrix} \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix} \quad (14-4)$$

where E_θ and E_ϕ are the spherical components in the θ - and ϕ -directions, ϕ is the azimuth angle, and ϕ_o is the reference polarization angle taken at the x-axis of the coordinate system. Axial ratio is the ratio of the major and minor axes of the polarization ellipse, usually on axis (or boresight) of the horn. If the polarization is perfectly circular, the axial ratio is 1. Axial ratio is usually expressed in dB.

Some horns have partially axisymmetric patterns, and this occurs when the far-field functions are $E_\theta(\theta, \phi) = F(\theta) \cos \phi$ and $E_\phi(\theta, \phi) = -G(\theta) \sin \phi$. In this case, $|F(\theta)|$ is the E -plane pattern function and $|G(\theta)|$ is the corresponding pattern in the H -plane. If $F = G$, the pattern is fully axisymmetric and, additionally, the cross-polarization is zero everywhere. Ideal corrugated horns have this property at the *balanced-hybrid condition* (see Section 14.4). Dominant mode circular horns have partially axisymmetric field functions. Horns with partial axisymmetry have zero cross-polarization in the E - and H -planes and peak cross-polarization in the intercardinal planes, the level given by the difference of the E - and H -plane pattern functions, and are $|F(\theta) - G(\theta)|/2$. The copolar pattern in the intercardinal planes is the average of the E - and H -plane pattern functions, $[F(\theta) + G(\theta)]/2$.

Input Match The input match is determined by two main reflection components: (1) reflection at the junction between the feeding waveguide and horn flare (i.e., the *throat*) and (2) reflection at the aperture due to transition from a guided wave to a radiating field.

The combination of these two components can produce a highly oscillatory return loss over the frequency band for some small horns, with a period approximately given by the distance between the junction and the aperture (i.e., period $\approx 720s/\lambda$ degrees where s is the distance between the throat and the aperture). Reflection from the transition to the flare is the most significant contributor for large aperture horns as the reflection at aperture diameters $> 2\lambda$ is usually small (> 20 dB). Most horns can be designed to have a return loss of 15 dB and higher. To achieve a good input match, considerable attention is paid to designing the input transition and the commencement of the horn flare.

Fabrication and Cost Horns are fabricated in many different ways from a variety of materials. The choice of fabrication method depends on cost and the intended application (weight, strength, heat, etc.). However, to achieve good efficiencies and performance, the inside metallic surfaces should be a good conductor (conductivity $\sigma > 1 \times 10^7$ S/m), unless the purpose is to intentionally load the horn. A lower conductor material is sometimes used for weight (e.g. carbon fiber) or mass-production (plastic or low-temperature metal with good casting properties) considerations, followed by a coating of high-conductor material (such as copper, gold, or silver) at least five skin-depths thick (i.e., $> 80 / \sqrt{f\sigma}$ mm, where f is the frequency in GHz). Additional passivation may be necessary to prevent corrosion or oxidation. Both material and labor costs are significant for this method of fabrication. In single applications, good performance can be obtained for moderate cost when horns are machined (on a lathe or machining center) from aluminum or jigged and brazed together from sheet-metal pieces, such as copper. Horns made from aluminum are dip-brazed while electroforming from copper is also possible. The material cost is negligible compared with labor costs in these approaches.

Effect of Horn Flare

Flaring a waveguide provides a directive radiation pattern and a smooth transition from the input to free-space. The field in the flare changes from a plane wavefront to a curved one, which is desirable for radiation. The surface across the face of the horn is called the *aperture*, and this surface is a convenient reference for calculating the radiated fields. The radiation characteristics of horns with aperture dimensions greater than about one wavelength may be calculated with reasonable accuracy by using equivalent currents (Huygens' principle) and an approximation to the aperture field. For single dominant-mode horns, the aperture field is approximated by the transverse electric field of that mode as it would exist in a uniform waveguide of the same cross-section and size as the horn mouth. In a waveguide that flares to a horn, it is clear that the boundary conditions require the existence of higher-order modes in addition to the fundamental mode. As a first approximation, it is usual to assume that the flare angle is small enough to permit the higher-mode field components to be neglected in comparison with those of the dominant mode.

Even when the flare angle is not particularly small, the above treatment gives useful design information if the effect of the curvature of the dominant-mode wavefront in the aperture is taken into account. In Figure 14-2, the field lines are shown as though the dominant mode were expanding as a cylindrical wave (in a sectoral horn) or a spherical wave (in a conical horn) with a radius of curvature at the aperture equal to L , the axial length of the horn in the section shown. The path difference between this curved wavefront and an ideal plane wave in the aperture constitutes a phase error that requires the planar fields of the dominant mode be modified by the inclusion of an approximate quadratic phase factor,

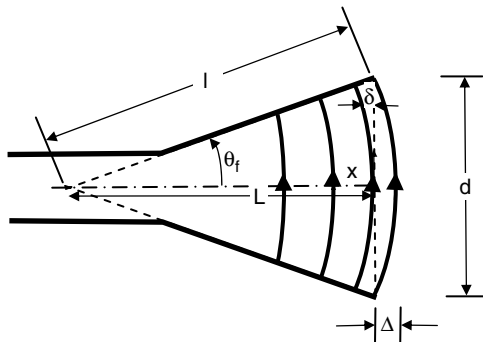


FIGURE 14-2 Phase error due to wavefront curvature in the horn aperture

$\exp(-jk\delta(x))$, where x is the normalized aperture coordinate and $k = 2\pi/\lambda$. In Figure 14-2, the distance from the wavefront to the aperture δ is

$$\delta(x) = \sqrt{L^2 + x^2} - L \approx \frac{x^2}{2L}; \quad 0 \leq |x| \leq \pm d/2 \quad (14-5)$$

Sometimes the slant length ℓ is employed instead of L in the quadratic phase factor. Throughout this chapter, the axial length L is used. At the aperture, the maximum excursion of the wavefront from planar is

$$\Delta = \frac{1}{2}d \tan \frac{\theta_f}{2} = 2\ell \sin^2 \frac{\theta_f}{2} \approx \frac{d^2}{8L} \quad (14-6)$$

where θ_f is the semi-flare angle of the taper. The approximation in Eq. 14-5 is reasonably accurate for $\theta_f < 35^\circ$; for larger angles the exact formula should be used.

The consequences of the quadratic phase factor are loss in gain accompanied by broadening of the radiation pattern, an increase in sidelobe levels, and filling in of the nulls. In the rest of this chapter, all pattern calculations are based on the above approximate model with quadratic-error correction applied as necessary.

Obliquity Factors

Radiation patterns of dominant (i.e. TE) mode horns are different in the principal planes because the transverse-field distribution and the associated field factor decays differently in the two orthogonal directions. The additional field decay is called an *obliquity*, or a *Huygens factor*, and is a pattern function multiplier that takes into account the effects of the mode propagation constant, mismatch at the horn aperture, and diffraction at the rim and surrounding surfaces.

On the θ -function in the far-field (E-plane for $\phi_o = 0$ in Eq. 14-4), the obliquity factor is represented by

$$F_\theta = \frac{1}{2} \left[1 + \frac{\beta}{k} \cos \theta + \Gamma \left(1 - \frac{\beta}{k} \cos \theta \right) \right] \quad (14-7)$$

while on the ϕ -function in the far-field (H -plane for $\phi_o = \pm 90$ degrees) the factor is

$$F_\phi = \frac{1}{2} \left[\cos \theta + \frac{\beta}{k} + \Gamma \left(\cos \theta - \frac{\beta}{k} \right) \right] \quad (14-8)$$

where β is the dominant mode propagation constant appropriate to the guide dimension at the horn mouth, Γ is the reflection coefficient presented by the aperture discontinuity and rim, while θ is the usual far-field polar angle. The first term corresponds to the incident mode and the second part its partial image scaled by Γ . In small-aperture horns, F_θ and F_ϕ are significantly different. For apertures larger than about two wavelengths, the aperture discontinuity is small, so that $\Gamma \rightarrow 0$, while $\beta/k \rightarrow 1$. Under these circumstances, F_θ and F_ϕ are nearly identical:

$$F_\theta \approx F_\phi \approx \frac{1}{2} (1 + \cos \theta) = \cos^2 \theta / 2 \quad (14-9)$$

When the horn aperture terminates in a metallic flange, the diffraction of the field at the aperture changes the obliquity factors to more general functions $F_\theta = f_1(\theta)$ and $F_\phi = f_2(\theta)$, where $f_1(\theta)$ and $f_2(\theta)$ depend on the geometry of the rim, the incident field, and the external currents. An extreme case is when the ground-plane is infinite, when $F_\theta = 1$ and $F_\phi = \cos \theta$. These obliquity factors will be used throughout this chapter for simplicity though recognizing that they can be replaced with Eq. 14-7 and Eq. 14-8 as the situation demands. Another extreme case is when the walls of the horn are infinitely thin. A closed-form solution for $f_1(\theta)$ and $f_2(\theta)$ is available for circular waveguides.¹¹⁰

Choice of Design Software

There are many excellent software packages available for analyzing horns and the choice can be bewildering. A wide variety of electromagnetic methods have been employed in these packages. Several methods and computer packages are referred to in this chapter and these are only a small sample of what is available to the antenna designer. However, every designer has their own special requirements. It is recommended, therefore, before investing a significant amount of time and money on a software package, consideration should be given to the most frequent design task. Choose aspects and trial problems to compare your application with each software package and assess their performance. These trial problems could be from a textbook or handbook such as this one. Evaluate the packages for accuracy, ease and speed of use, level of software support offered by the vendor, portability to other applications and to other computer platforms, and annual license fee. The effort expended in selecting a software package that best suits your needs will not go unrewarded in the long run.

14.2 RECTANGULAR HORNS

The first rectangular horns of the pyramidal variety were used by J. C. Bose in 1897, which he termed a “collecting funnel.” However, horns were not studied in detail until forty years later.¹ Their forerunner was, of course, the hollow rectangular waveguide

analyzed (for the first time) only a few years before the horn. Since the waveguide represents a useful approximation for rectangular horns, we commence this section with a brief overview of rectangular waveguide results.

Radiation from Rectangular Waveguide

An open-ended rectangular waveguide operating in the dominant mode (TE_{10} or TE_{01}) has a useful broad radiation pattern. When the aperture terminates in an infinite ground-plane the radiated far-fields by the TE_{10} mode are given by

$$E_{\theta}(r, \theta, \phi) = \frac{2abE_o}{\pi} \frac{e^{-jkr}}{r} \cos \phi S\left(U \frac{b}{2}\right) C\left(V \frac{a}{2}\right) \quad (14-10a)$$

$$E_{\phi}(r, \theta, \phi) = -\frac{2abE_o}{\pi} \frac{e^{-jkr}}{r} \sin \phi S\left(U \frac{b}{2}\right) C\left(V \frac{a}{2}\right) \cos \theta \quad (14-10b)$$

where

$$U = \frac{1}{\lambda} \sin \theta \cos \phi,$$

$$V = \frac{1}{\lambda} \sin \theta \sin \phi,$$

$$S(x) = \frac{\sin x}{x} \quad \text{and} \quad C(x) = \frac{\cos x}{1 - \left(\frac{2x}{\pi}\right)^2}.$$

The waveguide width is a (the dimension in the H -plane) and the height is b (the dimension in the E -plane). E_o is a constant. Above cutoff of the TE_{10} mode, the maximum gain of this antenna is

$$G_{\max} = \frac{32}{\pi} \frac{ab}{\lambda^2} \frac{k}{\beta} \quad (14-11)$$

where $\beta = \sqrt{k^2 - (\pi/a)^2}$ is the propagation constant of the mode. Well above cutoff $\beta/k \rightarrow 1$, in which case the aperture efficiency is $8/\pi^2$ or about 81%.

Sectoral Horns

Sectoral horns are produced when the rectangular waveguide input is flared in either the E - or H -plane. The rectangular horn shown in Figure 14-3 is the result that occurs when the waveguide is flared in both planes. The radiation patterns of sectoral horns may be approximately computed^{2,16} as described in "Effect of Horn Flare." The results of such calculations in the planes of the flare are shown in Figures 14-4 and 14-5 for E - and H -plane flares, respectively. In each case, there is a family of curves corresponding to the normalized cylindrical-wave error $\Delta_{e,h}/\lambda$ due to the flare, where

$$\Delta_e = \frac{b}{2} \tan \frac{\theta_e}{2} \quad \text{and} \quad \Delta_h = \frac{a}{2} \tan \frac{\theta_h}{2} \quad (14-12)$$

$\theta_{e,h}$ is the half-flare angle in the E - and H -plane, respectively.

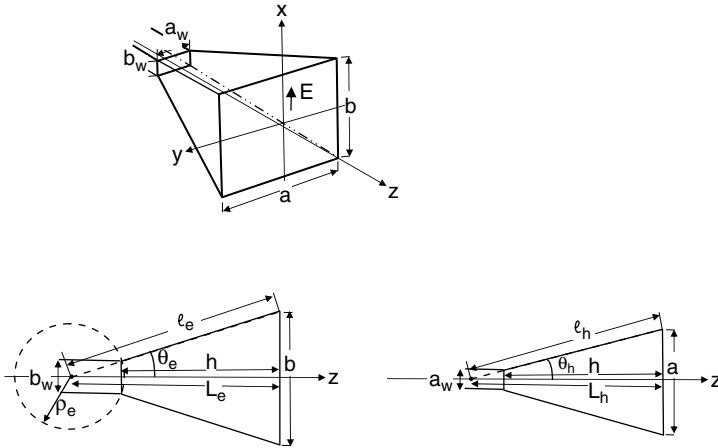


FIGURE 14-3 Geometry of rectangular horn: (a) overall perspective; (b) E -plane projection; (c) H -plane projection

For the same values of $\Delta_{e,h}/\lambda$, the phase error due to the flare has a much more pronounced effect in the E -plane than in the H -plane. The reason is simply that the field is constant across the aperture of a sectoral horn flared in the E -plane but varies sinusoidally (i.e. tapered) for one flared in the H -plane. The patterns also exhibit much higher minor-lobe levels for the former case.

The curves in both Figures 14-4 and 14-5 are reasonably accurate when either a or b is at least several wavelengths. For smaller apertures, it is necessary to multiply by the appropriate obliquity factor, F_θ or F_ϕ . The approximate form given by Eq. 14-9 is reasonably accurate for aperture dimensions greater than about 1.5λ .

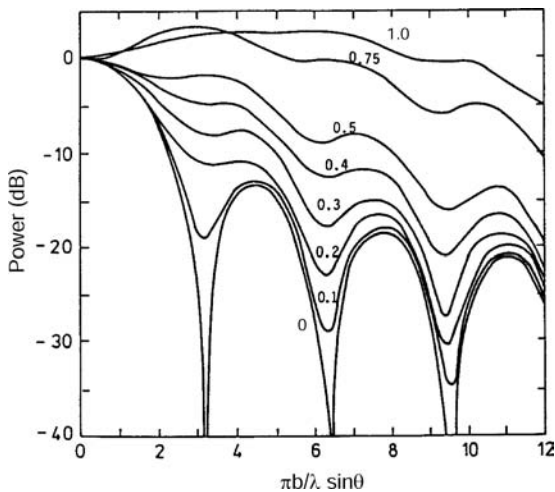


FIGURE 14-4 Universal radiation patterns of horns flared in the E -plane (sectoral or pyramidal). The parameter is the normalized phase error Δ_e/λ .

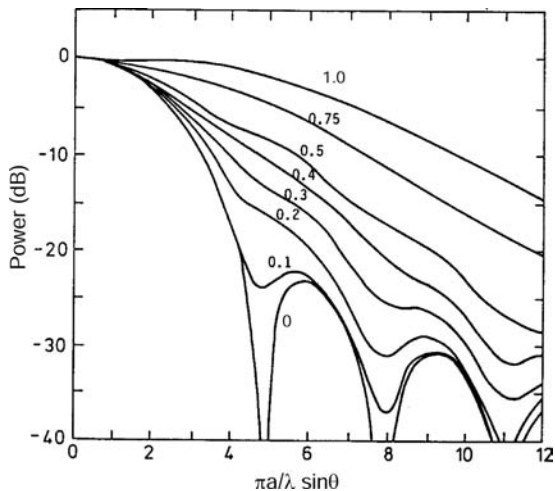


FIGURE 14-5 Universal radiation patterns of horns flared in the H -plane (sectoral or pyramidal). The parameter is the normalized phase error Δ_h/λ .

The aperture field method is incapable of predicting the radiation pattern in the rear sector of a horn antenna without the wall currents and is not accurate in the wide-angle sidelobe region because it neglects diffraction at the aperture edges. The prediction of wide-angle and rearward-radiated fields is possible by using the techniques of the geometrical and physical theories of diffraction (GTD and PTD). It is impractical to attempt such a treatment here; the interested reader should consult the references.^{2,111-115}

Pyramidal Horns

The pyramidal horn of Figure 14-3 flares in both the E - and the H -planes. Its radiation patterns in these planes are the same as those of the corresponding flared sectoral horns, so they are limiting cases of the pyramidal horn. Thus, the E -plane patterns of a pyramidal horn are as shown in Figure 14-4, while its H -plane patterns are those shown in Figure 14-5. The horns are assumed to be fed from a rectangular waveguide with dimensions in the H - and E -planes of a_w and b_w , respectively. The axial lengths in these planes are

$$L_h = ha/(a - a_w) \quad \text{and} \quad L_e = hb/(b - b_w) \quad (14-13)$$

where h is the length of the flared section. For aperture dimensions greater than about 1λ , the patterns in the two planes may be controlled independently by adjusting the mouth dimensions.

The electric field in the far zone of the aperture of a pyramidal horn located in a ground-plane and containing the TE_{mn} mode is given by

$$(E_\theta(r, \theta, \phi))_{mn} = A_{mn} \frac{\sqrt{L_e L_h}}{4} \frac{e^{-jk_r r}}{r} e^{-j\pi(Ua + Vb)} \left[\cos \phi (F_+(U) + F_-(U))(G_+(V) - G_-(V)) \frac{n}{b} \right. \\ \left. + \sin \phi (F_+(U) - F_-(U))(G_+(V) + G_-(V)) \frac{m}{a} \right] \quad (14-14a)$$

$$(E_\phi(r, \theta, \phi))_{mn} = -A_{mn} \frac{\sqrt{L_e L_h}}{4} \frac{e^{-jkr}}{r} e^{-j\pi(Ua+Vb)} \cos\theta \left[\sin\phi (F_+(U) + F_-(U))(G_+(V) - G_-(V)) \frac{n}{b} \right. \\ \left. - \cos\phi (F_+(U) - F_-(U))(G_+(V) + G_-(V)) \frac{m}{a} \right] \quad (14-14b)$$

where

$$F_\pm(U) = e^{j\alpha_\pm(a+\alpha_\pm)\frac{k}{2L_h}} \left\{ K\left(\sqrt{\frac{k}{2L_h}}\left(\frac{a}{2} - \alpha_\pm\right)\right) + K\left(\sqrt{\frac{k}{2L_h}}\left(\frac{a}{2} + \alpha_\pm\right)\right) \right\} \\ G_\pm(V) = e^{j\beta_\pm(b+\beta_\pm)\frac{k}{2L_e}} \left\{ K\left(\sqrt{\frac{k}{2L_e}}\left(\frac{b}{2} - \beta_\pm\right)\right) + K\left(\sqrt{\frac{k}{2L_e}}\left(\frac{b}{2} + \beta_\pm\right)\right) \right\} \\ \alpha_\pm = \left(2U \pm \frac{m}{a}\right) \frac{\pi L_h}{k}, \beta_\pm = \left(2V \pm \frac{n}{b}\right) \frac{\pi L_e}{k}$$

$U = 1/\lambda \sin\theta \cos\phi$, $V = 1/\lambda \sin\theta \sin\phi$, L_e , and L_h are the axial lengths from the apex to the aperture in the E - and H -planes, while $K(z) = \int_0^z \exp(-j\xi^2) d\xi$ is the complex Fresnel integral, which can be integrated numerically or calculated approximately.¹¹⁶ The electric fields of the TM_{mn} modes are similar, and these are obtained by reversing the sign on the K -function in the parentheses (Eq. 14-14a) from $+$ to $-$ and similarly in Eq. 14-14b from $-$ to $+$. The E - or H -plane sectoral horns are special cases obtained by allowing L_h or $L_e \rightarrow \infty$ in Eq. 14-14, respectively. The distance of the phase center from the aperture normalized to the axial length in the E - and H -planes of pyramidal (and sectoral) horns, which was calculated using Eq. 14-3 and the phase of Eq. 14-14 for the main beam between the -12 dB points, are shown in Figure 14-6 as a function of $\Delta_{e,h}/\lambda$.

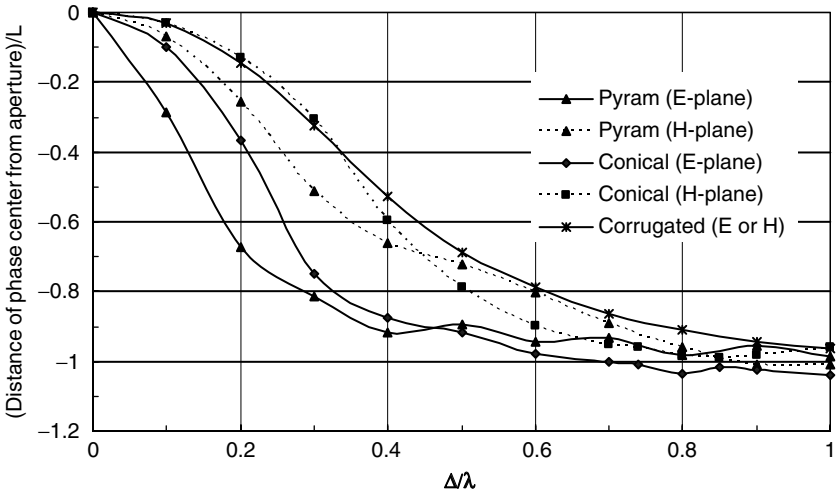


FIGURE 14-6 Phase center relative to the aperture in the planes of various horns as a function of the parameter Δ/λ , where Δ is the maximum deviation of the wavefront from planar in the designated plane. The minus sign indicates the phase center is located inside the horn.

The design of a pyramidal horn for specified beamwidths in the principal planes must be undertaken iteratively because of the need to fit the throat of the horn to a given input waveguide. A design procedure is outlined here:

1. Starting with the E -plane pattern, select a trial value for b to give an approximation to the desired half-beamwidth θ_{ebw} in degrees. Estimate b/λ from Figure 14-4 or use $b/\lambda \approx 44/\theta_{\text{ebw}}$. Select a value of horn length h , and calculate L_e from Eq. 14-13.
2. Compute the E -plane radiation pattern using Eq. 14-14. Refine b and h to create the desired beamwidth.
3. Repeat this procedure for the H -plane pattern. Choose an initial dimension a/λ from Figure 14-5 or use $a/\lambda \approx 40/(\theta_{\text{hbw}} - 15)$ where θ_{hbw} is the desired half-beamwidth in the H -plane in degrees. Calculate L_h from Eq. 14-13.
4. Compute and refine the H -plane radiation pattern using Eq. 14-14. Vary h , a , and b as needed to improve the design.
5. Once satisfied with this approximate design, refine further to achieve the desired performance (i.e., input match, polarization, etc.) by simulating the structure with accurate design software.

Gain

The maximum absolute gain of a rectangular horn radiating predominantly in the $\text{TE}_{10}\backslash\text{TE}_{01}$ mode is

$$G_{\text{max}} = \frac{8}{\pi} \frac{a}{b} r^2 \left(|E_{\theta}(r, 0, 0)|^2 + |E_{\phi}(r, 0, 0)|^2 \right) \bigg|_{\substack{m=1, n=0 \\ m=0, n=1}} \quad (14-15)$$

where the electric field components E_{θ} and E_{ϕ} are given by Eq. 14-14 and $k \gg \pi/a$.

The gain of sectoral and pyramidal horns may be determined from Eq. 14-15 or from the normalized gain curves of Figures 14-7 and 14-8, provided the narrow dimension is at

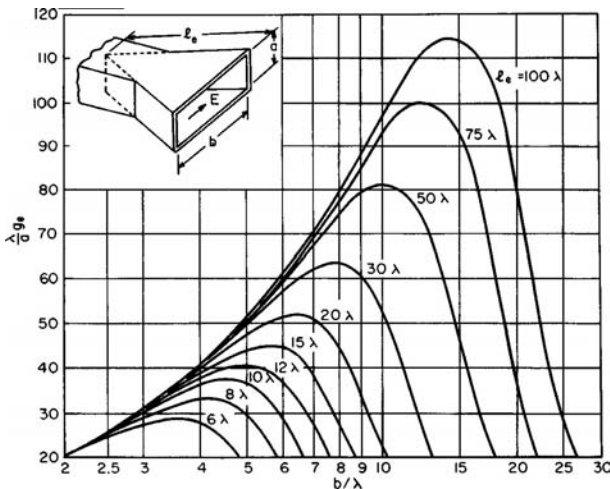


FIGURE 14-7 Normalized gain of E -plane sectoral horn ($a \leq \lambda$)

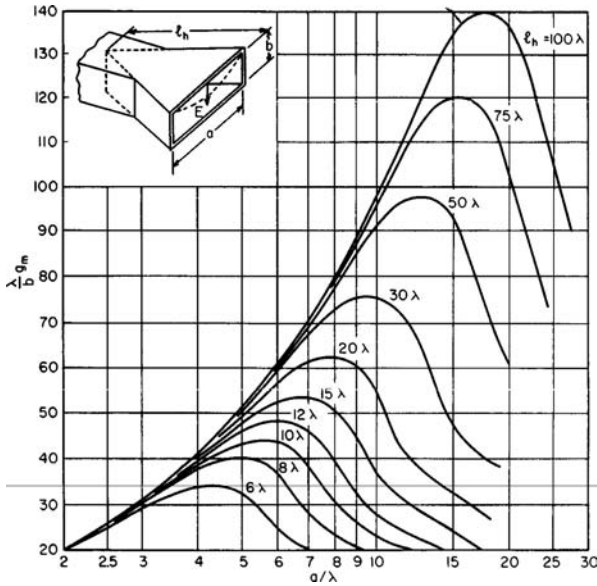


FIGURE 14-8 Normalized gain of H -plane sectoral horn ($b \leq \lambda$)

least one wavelength. The parameters g_e and g_h are the gains of E - and H -plane sectoral horns ($g_{e,h}$ has no dimension). It is seen that the gain initially increases as the aperture dimension increases, and then reaches a maximum before decreasing. The decrease in gain is due to increasing cylindrical-wave phase error as the aperture increases while the length remains constant.

The points of maximum gain in each case occur approximately where the phase errors at the aperture in the E - and H -planes are respectively 90° and 135° . That is

$$k\Delta_e \approx \pi b^2 / (4\lambda L_e) = \frac{\pi}{2} \quad (14-16)$$

$$k\Delta_h \approx \pi a^2 / (4\lambda L_h) = \frac{3\pi}{4} \quad (14-17)$$

The gain of both sectoral and pyramidal horns¹¹⁷ may be expressed by

$$G = G_0 R_e R_h \quad (14-18)$$

where the gain G_0 relative to an isotropic radiator is given by

$$G_0 = 32ab / (\pi\lambda^2) \quad (14-19)$$

and $R_{e,h}$ is gain reduction factors equal to or less than unity, which account for the loss in gain due to phase error caused by the horn flare; thus $R_{e,h}$ is a function of $\Delta_{e,h}/\lambda$. Both factors can be computed from Eq. 14-15. They are shown graphically in Figure 14-9 and are listed in Table 14-2.¹¹⁸ For a sectoral horn, one of the axial lengths will be infinite, and the corresponding gain reduction factor $R_{e,h}$ will be 0 dB. An alternative way of calculating the gain of a pyramidal horn is to take the product of the gains of the corresponding sectoral horns, obtained from Figures 14-7 and 14-8, and then to multiply by the factor $\pi/32$.

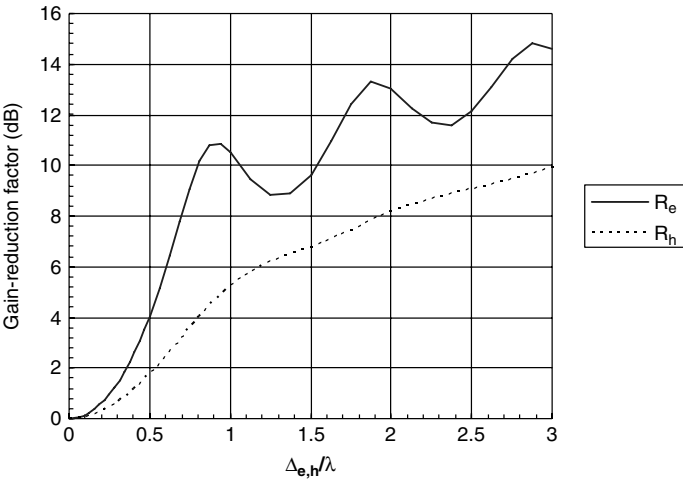


FIGURE 14-9 Gain-reduction factor for *E*- and *H*-plane flares

TABLE 14-2 Gain-reduction Factors $R_{e,h}$ for Sectoral and Pyramidal Horns¹¹⁸

$16\Delta_{e,h}/\lambda$	R_e , dB	R_h , dB	$16\Delta_{e,h}/\lambda$	R_e , dB	R_h , dB
1.0	0.060	0.029	14.0	10.783	4.486
1.5	0.134	0.064	15.0	10.849	4.892
2.0	0.239	0.114	16.0	10.502	5.252
2.5	0.374	0.179	18.0	9.474	5.819
3.0	0.541	0.257	20.0	8.847	6.210
3.5	0.738	0.349	22.0	8.901	6.504
4.0	0.967	0.454	24.0	9.637	6.785
4.5	1.229	0.573	26.0	10.938	7.102
5.0	1.525	0.705	28.0	12.430	7.460
5.5	1.854	0.850	30.0	13.312	7.831
6.0	2.218	1.007	32.0	13.052	8.175
6.5	2.618	1.176	34.0	12.251	8.460
7.0	3.054	1.357	36.0	11.666	8.684
7.5	3.527	1.547	38.0	11.607	8.869
8.0	4.037	1.748	40.0	12.121	9.047
9.0	5.166	2.175	42.0	13.104	9.243
10.0	6.427	2.630	44.0	14.221	9.462
11.0	7.769	3.101	46.0	14.851	9.692
12.0	9.081	3.577	48.0	14.619	9.911
13.0	10.163	4.043	50.0	13.937	10.101

In the previous discussion of gain, far-field conditions are implicitly understood. It may be shown,¹¹⁸ however, that Eqs. 14-15 and 14-18 may be used at a finite range (in the Fresnel zone) by a simple redefinition of terms. This is possible because the effect of finite range is to increase slightly the already existing quadratic phase error in the aperture due to the flare. Define new effective axial lengths as follows:

$$1/L'_e = 1/L_e + 1/r \quad \text{and} \quad 1/L'_h = 1/L_h + 1/r \quad (14-20)$$

where r is the distance from the aperture center to the field point. Then for all equations already defined for gain (Eqs. 14-15 to 14-19), G , R_e , and R_h are determined as before, but the effective lengths defined by Eq. 14-20 are used instead—even for sectoral horns where either L_e or $L_h \rightarrow \infty$.

When the a dimension (E -plane) in a sectoral or pyramidal horn is small $a \ll \lambda$, the close proximity of the magnetic walls causes errors in the gain equations (Eqs. 14-18 and 14-20). One can use the accurate expression (Eq. 14-15) or the approximation¹¹⁹:

$$G = R_e R_h \frac{4k^2 ab}{\pi^2 (1 + k/\beta)} \exp \left[\frac{ka}{2} \left(1 - \frac{\beta}{k} \right) \right] \quad (14-21)$$

in which R_e and R_h are as before except that in calculating R_e the parameter Δ_e must be slightly modified. Thus

$$\Delta_e = \beta b^2 / \left(8kL_e \cos^2 \frac{\theta_e}{2} \right) \quad (14-22)$$

where θ_e is the semi-flare angle in the E -plane and β is the phase shift of the dominant mode appropriate to the dimension a .

Equation 14-18, which may be called *Schelkunoff's gain formula*, includes the geometrical-optics field in the aperture of the horn and the effects of singly diffracted fields at the aperture edges. It does not, however, include effects due to multiple edge diffractions or diffracted fields that are reflected from interior horn walls. These effects manifest themselves in the form of small oscillations about the monotonic curve predicted by Schelkunoff's equations. The oscillations can be of the order of ± 0.5 dB for pyramidal horns with a gain of about 12 dB. They decrease to about ± 0.2 dB for horns with a gain of about 18 dB, and they appear to be less than ± 0.1 dB when the gain exceeds 23 dB.¹²⁰⁻¹²¹

Optimum-Gain Horns

A class of rectangular horns that historically has been used widely as a gain reference is the optimum-gain horn (sometimes called a "standard" gain horn). This is a pyramidal horn with dimensions such that maximum far-field gain is produced for given axial lengths in the E - and H -planes. From what has been said in connection with Eqs. 14-16 and 14-17, an optimum-gain horn will have aperture dimensions given approximately by

$$a = \sqrt{\alpha_1 L_h} \quad b = \sqrt{\beta_1 L_e} \quad (14-23)$$

where $\alpha_1 \approx 3\lambda$ and $\beta_1 \approx 2\lambda$. The effective gain of an optimum-gain horn is very close to 50% of its actual aperture area; its gain is

$$\text{Gain (dB)} = 8.1 + 10 \log_{10} (ab/\lambda^2) \quad (14-24)$$

Such a horn will have semi-half-power beamwidths (HPBW) given approximately by

$$\theta_{\text{ebw}} \approx 60 b \sin(0.45\lambda/b) \text{ degrees,} \quad E\text{-plane}$$

$$\theta_{\text{hbw}} \approx 60 a \sin(0.63\lambda/a) \text{ degrees,} \quad H\text{-plane}$$

and first-sidelobe levels of -9 dB and -16 dB, respectively. The H -plane sidelobe is not well defined, amounting to no more than a shoulder on the main beam.

In designing an optimum-gain horn, one generally knows the frequency, the desired gain G_d (initially given in dB, but converted to its value for the calculations), and the dimensions of the feed waveguide a_w and b_w .

1. The horn geometry needs to satisfy the following geometric constraint to be physically realizable (from Eq. 14-13, h is the same in both planes):

$$L_e \left(1 - \frac{b_w}{b} \right) = L_h \left(1 - \frac{a_w}{a} \right) \quad (14-25)$$

This equation results in a quartic polynomial for which the only solution of interest (real solution and $0 < a_w < a$) is

$$a = \sqrt{A_1 + A_2} - \frac{\left(b_w c_o - \frac{a_w^2}{8} \right)}{4A_1} + \frac{a_w}{4} \quad (14-26)$$

where

$$c_o = \frac{(G_d \lambda^2)}{g_1} \frac{\alpha_1}{\beta_1},$$

$$A_1 = \sqrt{A_2^2 + 3 \left(u + \frac{P}{u} \right)^2},$$

$$A_2 = \left(u - \frac{P}{u} \right) + \frac{a_w^2}{8},$$

$$u = \left| Q + \sqrt{Q^2 + P^3} \right|^{\frac{1}{3}},$$

$$P = \frac{c_o}{12} \left(\frac{(G_d \lambda^2)}{g_1} - \frac{a_w b_w}{4} \right), \quad \text{and} \quad Q = \frac{c_o^2}{128} \left(a_w^2 \frac{\beta_1}{\alpha_1} - b_w^2 \right).$$

For a pyramidal horn, $g_1 = 2\pi$, $\alpha_1 = 3\lambda$, and $\beta_1 = 2\lambda$.

2. Calculate the remaining horn parameters from $b = \frac{(G_d \lambda^2)}{g_1 a}$, $L_h = a^2 / \alpha_1$, $L_e = b^2 / \beta_1$, and $h = L_h (1 - a_w / a)$.
3. Refine the design with your most accurate software to take account of wall effects to obtain the desired gain and input match over the specified bandwidth.

A special case of Eq. 14-26 is when $Q = 0$. This gives $b_w = a_w \sqrt{\beta_1 / \alpha_1}$, so that $a = \lambda \sqrt{(G_d / g_1) (\alpha_1 / \beta_1)^{1/2}}$ and $b = \lambda \sqrt{(G_d / g_1) (\beta_1 / \alpha_1)^{1/2}}$, resulting in a simple dependence on desired gain and wavelength as in the case of conical horns (see Section 14.3).

Input Match

The input reflection coefficient of sectoral and pyramidal horns depends on the mismatch between the horn mouth and free-space, the length of the guide between the horn aperture and throat, and the mismatch at the junction between the uniform guide and the throat. Aperture mismatch can be significant in E -plane sectoral horns when b is large compared with the width a . The input return loss of an E -plane horn varies periodically as the horn length is varied, with minima occurring approximately every half-guide wavelength.^{16,122} The guide wavelength in the E -plane sectoral horn is equal to that in the feeding rectangular waveguide. The input reflection coefficient for an E -plane sectoral horn is approximately¹²³:

$$\Gamma_{in} \approx \Gamma_{je} + \Gamma_{ac} \exp(-2j\beta\ell_e) \quad (14-27)$$

where

$$\Gamma_{je} = \frac{1 - y_{je}}{1 + y_{je}} \quad (14-28)$$

is the reflection coefficient of the junction between the waveguide and the flare, $y_{je} = jH_0^{(2)}(k\rho_e)/H_0^{(2)}(k\rho_e)$, and ρ_e is the radius from the apex to the flare (see Figure 14-3b) and

$$\begin{aligned} \Gamma_{ac} = & -\exp\left(j2\Phi_e(k\rho_e) + ph\left\{H_0^{(1)}(k\rho_e)\right\}\right) \left(\frac{1}{4\theta_e\beta\ell_e} + \frac{1}{2\beta a} \left(\frac{\beta}{k} - 1\right)\right) \\ & \times \exp\left(\frac{1}{\beta} \sqrt{\frac{2k}{\pi a}} \exp\left(j\left(ka + \frac{\pi}{4}\right)\right)\right) \end{aligned} \quad (14-29)$$

is the aperture reflection coefficient, where $\Phi_e(x) = a \tan\left(\frac{\text{Re}\{Z_e(x)\}}{1 + \text{Im}\{Z_e(x)\}}\right)$, $Z_e(x) = \frac{H_0^{(1)}(x)}{H_1^{(1)}(x)}$, and θ_e is the semi-flare angle in the E -plane (see Figure 14-3b). In Eqs. 14-28 and 14-29, $H_n^{(1)}(x)$ and $H_n^{(2)}(x)$ are the Hankel functions of types 1 and 2 of order n . An example of the use of Eq. 14-27 is shown in Figure 14-10 for an E -plane sectoral horn with $a = 7.214$ cm, $b = 3.404$ cm, $\ell_e = 42.15$ cm, and $\theta_e = 16.5^\circ$. The approximate expression in Eq. 14-27 is in reasonable agreement with measured results.¹²³

In all dominant-mode rectangular horns, the reflection coefficient due to the aperture Γ_{ap} of the TE_{10} mode at the aperture is given approximately by

$$\Gamma_{ap} \approx \left(\frac{1 - y_{11}}{1 + y_{11}}\right) \quad (14-30)$$

where the aperture admittance y_{11} of this mode when input on an infinite flange¹²² is given by

$$\begin{aligned} y_{11} = & \frac{2jk^2}{\pi ab\beta} \left(\int_0^{\alpha_o} d\alpha \int_0^{a \sec \alpha} dx + \int_{\alpha_o}^{\pi/2} d\alpha \int_0^{b \cos \alpha} dx \right) \exp(-jkx)(b - x \sin \alpha) \\ & \times \left(\frac{1}{k_c} \left(1 + \left(\frac{k_c}{k} \right)^2 \right) \sin(k_c x \cos \alpha) + \left(1 - \left(\frac{k_c}{k} \right)^2 \right) (a - x \cos \alpha) \cos(k_c x \cos \alpha) \right) \end{aligned} \quad (14-31)$$

where $k_c = \pi/a$ and $\alpha_o = a \tan(b/a)$. The return loss for the aperture is typically better than 25 dB for $a > 2\lambda$ and $a < b < 2a$.

The mismatch of an H -plane sectoral horn differs somewhat from that of the E -plane horn, in that the mismatch at the junction of the horn and rectangular waveguide is much

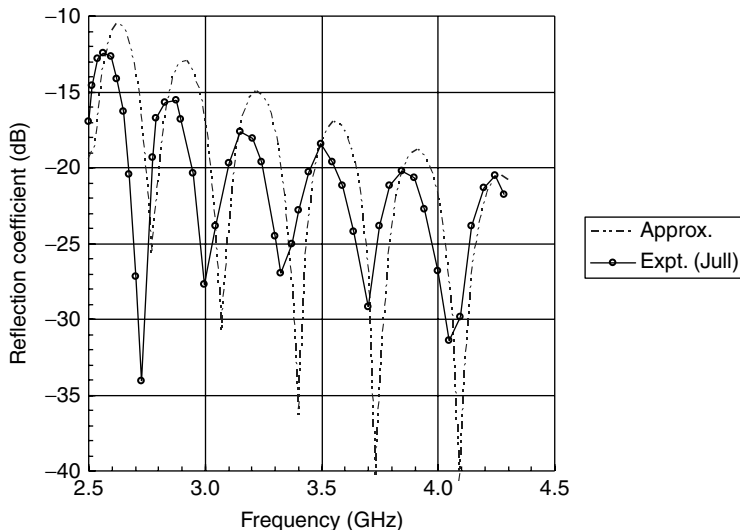


FIGURE 14-10 Reflection coefficient of E -plane sectoral horn ($a = 7.214$ cm, $b = 3.404$ cm, $\ell_e = 42.15$ cm, $\theta_e = 16.5^\circ$). The experimental data are from Jull.¹²³

smaller than that at the horn aperture. The corresponding equation to Eq. 14-28 for H -plane junctions is not very accurate, and it has been found that a good estimate of the mismatch can be obtained from aperture reflection only. Thus, for the H -plane sectoral horn $\Gamma_{in} \approx \Gamma_{ap} \exp(-2j\beta h)$, where Γ_{ap} is given by Eq. 14-31. The result is that the input match to an H -plane sectoral horn is almost constant in magnitude (for a constant aperture width a) and varies only in phase as the length of the horn is changed.¹⁶

A broadband match is best obtained in the E -plane sectoral horn by treating the two discontinuities separately. The mismatch at the junction of the waveguide and horn may be eliminated with a reactive window (usually inductive) at the junction. The horn aperture is frequently matched to free-space by using a radome of appropriate thickness and dielectric constant. If no radome is used, a match may be obtained by using small reactive discontinuities at the aperture.

The problem of obtaining a broadband match with the H -plane horn is somewhat easier because of the small discontinuity at the horn throat. Generally, only the aperture needs to be matched, and the techniques mentioned for the E -plane horn may be employed.

An estimate of the input reflection coefficient in a pyramidal horn can be found from

$$\Gamma_{in} \approx \Gamma_{je} + \Gamma_{ap} \exp(-2j\beta h). \quad (14-32)$$

where Γ_{je} is given by Eq. 14-28 and Γ_{ap} by Eq. 14-30. Results from Eq. 14-32 and mode matching (MM) software with an infinite flange are shown in Figure 14-11 along with experimental results for a 20 dB optimum-gain horn ($a = 100$ mm, $b = 74$ mm, $h = 217$ mm, WR-75 input) supplied by Flann Microwave (model 1724-20) (shown in Figure 14.1a). Also shown are results for this horn from the commercial software package CST Microwave Studio.¹²⁴ Differences between the accurate theory and experiment can be attributed to conditions at the aperture (i.e., thin or large flange) and the junction between the waveguide and the flare. Generally, horns having a gain of 20 dB or more and moderate flare angles, such as optimum horns, usually are well matched to a rectangular waveguide supporting

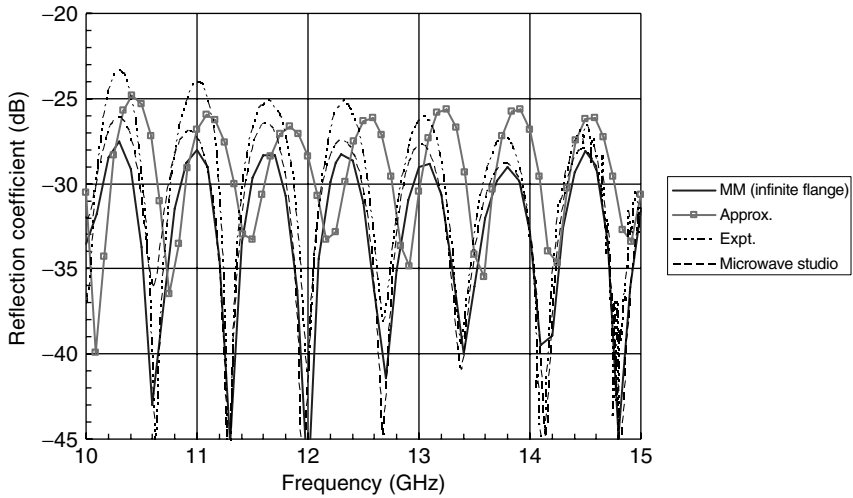


FIGURE 14-11 Reflection coefficient of Ku-band 20 dB optimum-gain horn with WR-75 input waveguide and $a = 100$ mm, $b = 74$ mm, and $h = 217$ mm

only the dominant mode. Where the overall match has to be improved, techniques similar to those described for sectoral horns may be adopted. However, when there are issues with matching, after an initial design matching section, steps in cross-section or irises can be introduced and assessed with accurate design software.

Other Rectangular Horns

There are a wide variety of other smooth-walled horns with a rectangular cross-section in general use. The most important ones are covered in this section.

Diagonal Horns The simplest example of a dual-mode horn⁸⁹ in which two spatially orthogonal dominant modes, the TE_{10} and the TE_{01} , are excited with equal amplitude and phase in a square waveguide that flares to a pyramidal horn. In this case, the electric field in the aperture is aligned with one of the diagonals. A simple transformation from rectangular guide to diagonal horn is shown in Figure 14-12. Another method of excitation is the use of coaxial-fed orthogonal probe couplers in square waveguides along with a suitable power splitter.

For apertures in which $d > 2\lambda$ the E -, H -, and intercardinal-plane patterns will be essentially identical to well below the 10-dB level, as is evident in the patterns shown in

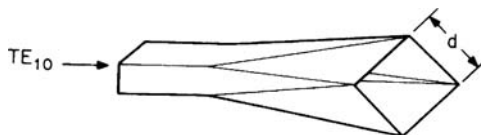


FIGURE 14-12 Transition from rectangular waveguide to diagonal horn

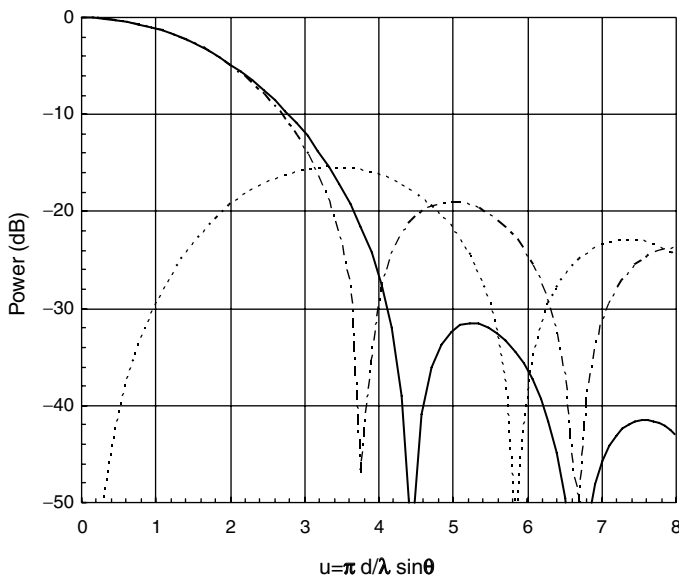


FIGURE 14-13 Principal- and intercardinal-plane patterns of narrow-flare angle diagonal horn ($d > 2\lambda$): —, E - and H -planes; ····, intercardinal copolar; and - · - ·, intercardinal, cross-polar

Figure 14-13 for narrow flare-angle ($< 10^\circ$). True axial symmetry is not achieved, however, because the intercardinal-plane patterns differ markedly in sidelobe structure from those in the principal planes. Moreover, a significant amount of cross-polarization is manifested in the intercardinal planes, as shown by the dashed curve in Figure 14-13.

By writing $u = \pi d \sin \theta / \lambda$, the principal patterns of the diagonal horn are given by

$$E_{e,h} = \pm F_{\theta,\phi} S(u/\sqrt{2}) \cdot C(u/\sqrt{2}) \quad (14-33)$$

and the cross-polarization is zero, while in the intercardinal planes

$$E_{cp, xp} = \frac{1}{2} (F_\theta \cdot S(u) \pm F_\phi \cdot C(u)) \quad (14-34)$$

where F_θ and F_ϕ are the obliquity factors discussed in the section, “Obliquity Factors,” earlier in the chapter, and the functions S and C are defined below Eq. 14-10. When $d > 2\lambda$, it was observed that $F_{\theta,\phi} \rightarrow 1$; this is the case for the patterns shown in Figure 14-13. Important characteristics of large-aperture diagonal horns are summarized in Table 14-3. Improved design accuracy for the radiation from medium flare angle ($< 35^\circ$) horns is obtained with Eq. 14-14 by combining equal amplitude TE_{10} ($m = 1, n = 0$) and TE_{01} ($m = 0, n = 1$) modes.

Prime-focus feed horns have small apertures ($d \sim \lambda$) for frequently used f/D ratios. The diagonal horn is at a disadvantage here because the disparity between the obliquity factors destroys the circular symmetry of the beam. It has been observed in practice, however, that in the neighborhood of $d = 1.23\lambda$, a diagonal horn possesses axial-beam symmetry with full widths of 40 and 73° at the -3 - and -10 -dB levels. Apparently, in this case, the

TABLE 14-3 Characteristics of Diagonal Horns⁸⁹

Parameter	Principal Planes	45° and 135° Planes
3-dB beamwidth,°	58.5 λ/d	58 λ/d
10-dB beamwidth,°	101 λ/d	98 λ/d
Angular position, first null,°	81 λ/d	70 λ/d
Angular position, first lobe,°	96 λ/d	92 λ/d
Level of first sidelobe	-31.5 dB	-19 dB
Angular position, second null,°	122 λ/d	122 λ/d
Angular position, second lobe,°	139 λ/d	147 λ/d
Level of second sidelobe	-41.5 dB	-24 dB

values of Γ and β/k (see “Obliquity Factors”) happen to be such that the obliquity factors are approximately equal.

Multimode, Stepped, and Profiled Rectangular Horns Multimoding in horns refers to the presence of more than one mode in the aperture at an intentionally significant level. Usually these modes are in addition to the dominant mode. There are essentially two main reasons for multimoding. One reason is to achieve the highest possible aperture efficiency for a given aperture size and the other is to lower the sidelobes in the E -plane to be similar to those in the H -plane. Both approaches will now be described.

The main method to date to design high-efficiency rectangular horns is based on achieving a uniform aperture distribution with unidirectional polarization. Horns designed this way employ TE_{0n} or TE_{n0} ($n = 1, 3, \dots$) modes with approximate amplitudes of $1/n$.^{16,20,14} Aperture efficiencies in excess of 90% can be achieved using this approach, compared with a maximum of about 81% with conventional rectangular horns. Steps are introduced along the length of the horn to excite the desired amplitude and a length of horn is used to bring the modes into phase equality. The bandwidth of operation at the highest efficiency is usually about 10%. However, the length of both conventional and dual-mode horns is relatively long due to the linear taper from the waveguide input to the waveguide step that is used to excite the TE_{0n}/TE_{n0} modes. Another way of achieving high aperture efficiency is to design a mode converter in the horn by stepping or profiling the cross-section in the flare to the aperture in order to excite modes of the desired amplitude and phase. It has been shown¹⁴ that the highest efficiency can be obtained by designing a mode converter that excites predominantly TE modes. However, when the horn is relatively small, there is significant reflection at the aperture. It has been shown that horns of width given approximately by

$$\frac{a}{\lambda} \approx \frac{n}{2} + \frac{(15+n)}{160} \quad (14-35)$$

where n is an odd integer (<9), the aperture efficiency is close to maximum, and $\sim 100\%$ over a modest bandwidth. Bird and Granet¹⁵ describe a way of achieving high efficiency by optimizing both profiles of the horns. An example of this type of horn is shown in Figure 14-1b.

Flare-angle changes may also be used in a pyramidal horn to generate the hybrid TE_{12}/TM_{12} pair, as shown by Cohn¹³ and illustrated in Figure 14-14. It is assumed that only the dominant mode propagates at the throat. Conversion to TE_{12}/TM_{12} modes

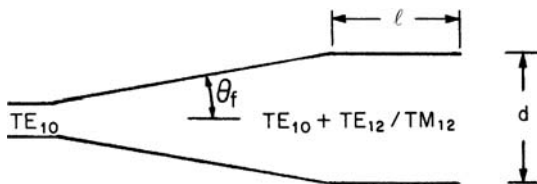


FIGURE 14-14 Section through rectangular horn with flare-angle change

occurs at the point where the flare-angle changes from θ_f to 0. Relative to unity for the TE_{10} mode, the TE_{12}/TM_{12} amplitude is

$$A \approx j2 \theta_f d / 3\lambda_g \quad (14-36)$$

providing θ_f is small (i.e., $< 23^\circ$), and where λ_g is the TE_{10} -guide wavelength corresponding to the waveguide dimension d . It is seen that the hybrid-mode phase leads that of the dominant mode by 90° . Therefore, the constant-diameter section ℓ must create a 270° differential shift to obtain correct phasing in the aperture. The length ℓ can be calculated from

$$\ell(\beta_{10} - \beta_{12}) = 3\pi/2 \quad (14-37)$$

where $\beta_{10} = k\sqrt{1 - (\lambda/2d)^2}$ and $\beta_{12} = k\sqrt{1 - (\lambda/2d)^2(4 + (d/w)^2)}$ where w is the other waveguide dimension (into the page of Figure 14-14). If no phasing section is used (i.e., $\ell = 0$), mode conversion still occurs and Eq. 14-36 applies as before. This means that the TE_{12} and TM_{12} energy is improperly phased relative to the TE_{10} and performance is that of a simple flared horn with spherical-wave error in the aperture. This complication must be included in any design.

Beam-shaping with a square aperture pyramidal horn can be accomplished by combining the dominant TE_{10} mode with a hybrid mixture of higher-order TE_{12} and TM_{12} modes, as shown by Jensen^{1,17} (Figure 14-15). These two higher-order modes, which have the same propagation constant, can exist as a hybrid pair, and their relative amplitudes can be adjusted to give a purely linearly polarized aperture field. When added in the proper phase to the TE_{10} field in the aperture, the resulting distribution is tapered in the E - as well as in the H -plane. If one assumes that dual-polarization capability is required, the input waveguide in Figure 14-15 will be square with a side of dimension d_1 such that the TE_{12}/TM_{12} pair cannot propagate; i.e., $d_1 < \sqrt{5}\lambda/2$. At the mode-generating step, both dimensions of the guide increase to d_2 , where $d_2 > \sqrt{5}\lambda/2$. The ratio d_2/d_1 is in the range 1.2 to 1.3, which means that d_1 is large enough to support 11, 02, and 20 index modes as well as the dominant TE_{10} (or TE_{01}). For this reason, a tapered transition is needed to transform from purely dominant-mode guide (for which $\lambda < 2d_0 < \sqrt{2}\lambda$) to the input guide of side d_1 . A constant-dimension phasing section of length ℓ (see Figure 14-14) is used to ensure correct phasing at the aperture. Unfortunately, the operation is relatively narrowband; 3 to 4% is typical.

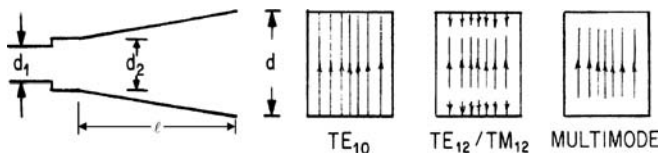


FIGURE 14-15 Dual-mode pyramidal horn

If the aperture is large enough to make the obliquity factors essentially unity ($d > 2\lambda$) and the semi-flare angle is small enough to render the spherical-wave error negligible, the patterns of a multimode horn are given by

$$\left. \begin{array}{ll} E\text{-plane} & E = S(u) \left(1 + \alpha\sqrt{2} \frac{u^2}{\pi^2 - u^2} \right) \\ H\text{-plane} & E = C(u) \end{array} \right\} \quad (14-38)$$

where S and C are defined beneath Eq. 14-10, $u = \pi d \sin \theta / \lambda$, and α represents the fractional TE_{12}/TM_{12} -mode content relative to unity for the TE_{10} mode. The interesting range for α is between 0.42 and 0.55, corresponding to a power level in the TE_{12}/TM_{12} mode between 7.5 and 5 dB below that in the dominant mode. The patterns shown in Figure 14-16 were calculated from Eq. 14-38 for $\alpha = 0.464$. Note that the main-beam shapes are nearly identical in the two planes down to the -20 -dB level, while the first E -plane sidelobe level is very low, approximately -45 dB. The cross-polar level in the 45° -plane is < -30 dB across most of the main beam. The full HPBW in all planes is $69\lambda/d$ degrees.

To design a horn using flare angle changes to give the patterns of Figure 14-16, it is required that $|A| = \alpha\sqrt{2}$ with $\alpha = 0.464$. Thus the condition is $2\theta_f d / 3\lambda_g = 0.656$. This gives $\theta_f = 56\lambda_g/d$ degrees. For a small-aperture horn (e.g., $d < 2\lambda$), this condition leads to a relatively large value for θ_f , which may violate the small-angle condition used in deriving Eq. 14-36. In this case, two or more separate flare changes may be used, since the A coefficients are additive, although frequency sensitivity will increase. Good pattern performance over a 30% bandwidth has been reported for the single-flare-change case.

Dielectric-lined Rectangular Horns Dielectric can be placed on the walls of a rectangular horn to increase the gain or aperture efficiency instead of multimoding (see Figure 14-17).¹¹⁻¹² The dielectric loading provides an extra degree of freedom in design, and in shaped-beam applications from satellites, a smaller number of elements can be used as the aperture efficiency is higher than conventional rectangular horns. The disadvantages are that horns

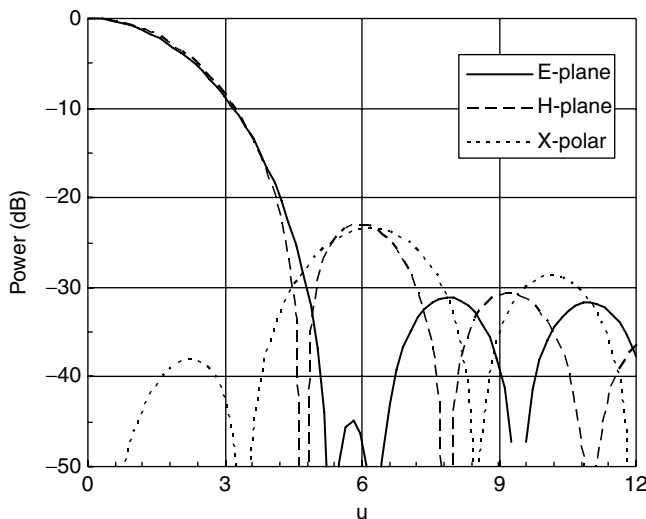


FIGURE 14-16 Calculated power patterns of a multimode pyramidal horn for $\alpha = 0.464$

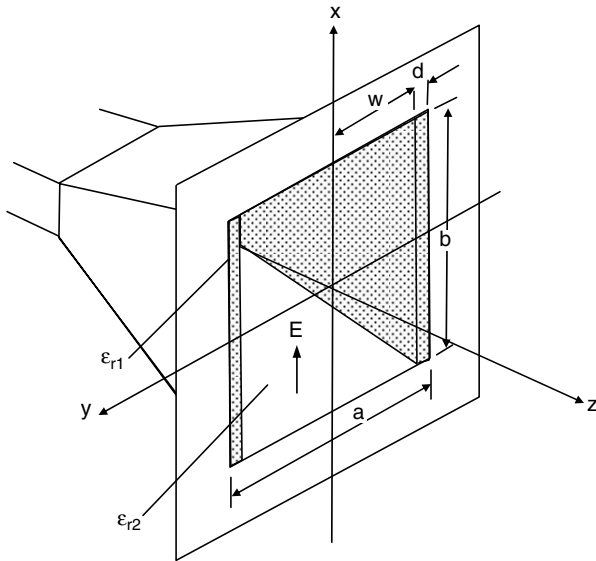


FIGURE 14-17 Geometry of dielectric-loaded rectangular horn with a flange

are heavier, have higher loss, are more difficult to fabricate (e.g., the dielectric must be fixed in close contact with the E -plane walls), and in space applications the material may outgas. The aim is to make the E - and H -plane sidelobes similar, and this is done by increasing the H -plane sidelobes. When a dielectric of thickness t is used, where

$$t \approx \frac{\lambda}{4\sqrt{\epsilon_{r1} - \epsilon_{r2}}} \quad (14-39)$$

and ϵ_{r1} and ϵ_{r2} are the dielectric constants in regions 1 and 2 shown in Figure 14-17, the field in the H -plane is also almost uniform. This improves the illumination efficiency of small and large aperture horns and is not limited by aperture size as in multimode horns.

The radiation fields can be obtained in a manner similar to that for the rectangular horn by assuming a quadratic phase factor on the dominant mode of a dielectric-loaded waveguide, the LSE_{10} mode. The result is an extension of Eq. 14-14 (with $m = 1$ and $n = 0$), in other words,

$$E_{\theta}(r, \theta, \phi) = A_{10} \frac{jk}{2\pi} \frac{e^{-jkr}}{r} N_x \cos \phi \quad (14-40a)$$

$$E_{\phi}(r, \theta, \phi) = -A_{10} \frac{jk}{2\pi} \frac{e^{-jkr}}{r} \cos \theta N_x \sin \phi \quad (14-40b)$$

where

$$N_x = \frac{\sqrt{L_e L_h}}{2\beta} e^{-j\pi Vb} [G_+(V) + G_-(V)] [F_{2+}(U) + F_{2+}(-U) + F_{2-}(U) + F_{2-}(-U) - j c_o (e^{j\pi Ua} (F_{1+}(-U) - F_{1+}(-U)) + e^{-j\pi Ua} (F_{1+}(U) - F_{1+}(U)))] \quad (14-41)$$

and where

$$F_{i\pm}(U) = e^{j\frac{\beta}{2L_h}\alpha_{i\pm}[a(2-i)+\alpha_{i\pm}]} \left\{ K\left(\sqrt{\frac{\beta}{2L_h}}(w-\alpha_{i\pm})\right) + K\left(\sqrt{\frac{\beta}{2L_h}}\left(\frac{a}{2}+\alpha_{i\pm}\right)\right) \right\}$$

and $G_{\pm}(V)$, $K(\cdot)$, U , and V are defined with Eq. 14-14, $\alpha_{i\pm} = (2\pi U \pm k_i) \frac{L_h}{\beta}$, $c_o = \frac{\cosh(k_2 w)}{\sin(k_1 d)}$, $k_i = \sqrt{k^2 \epsilon_{ri} - \beta^2}$, and β is the propagation constant of the LSE₁₀ mode that is obtained by solving $k_2 \tan k_2 w = k_1 \cot k_1 d$. An approximate lower bound for β is

$$\frac{\beta}{k} \approx \sqrt{1 + (\epsilon_{r1} - 1) \left(\frac{2d}{a} - \frac{1}{\pi} \sin\left(\frac{2\pi d}{a}\right) \right) - \left(\frac{\pi}{a} \right)^2 + (\epsilon_{r2} - 1) \frac{2w}{a} \left(1 - \frac{1}{\pi} \sin\left(\frac{2\pi w}{a}\right) \right)} \quad (14-42)$$

which can be used as an initial value in finding accurately the roots of the transcendental equation. Note that the argument of the Fresnel integral $K(z)$ in region 1 ($F_{1\pm}$) is complex since k_1 is imaginary for high-efficiency horns. Padé approximations to these integrals are available.¹¹⁶ Computed and measured radiation patterns in the H -plane of a dielectric-loaded horn are shown in Figure 14-18 at 12.5 GHz. The dimensions of the horn in the aperture are $a = 61$ mm, $b = 52.8$ mm, $d = t = 4.1$ mm, $\epsilon_{r1} = 3.07$, $\epsilon_{r2} = 1$, $h = 228.6$ mm, and it was fed from a WR75 waveguide. The theoretical values were obtained from Eq. 14-40. The sidelobe levels in the H -plane are about -16 dB down, and this compares with about -13.5 dB (measured) for the E -plane. However, the main beam is very similar.

The absolute gain of the dielectric-loaded horn is given by

$$G_{\max} = \frac{2\pi}{P_{\text{in}}} r^2 (|E_{\theta}(r, 0, 0)|^2 + |E_{\phi}(r, 0, 0)|^2) \quad (14-43)$$

where $P_{\text{in}} = \sqrt{\frac{\beta}{k}} \frac{b}{2} [w(1 + S(2k_2 w) + d |c_o|^2 (1 - S(2k_1 d))]$, S is defined below Eq. 14-10, and the fields are given by Eq. 14-40. The aperture efficiency of a dielectric-loaded horn can exceed that of the unloaded horn providing $a > \lambda$. Absolute gain results are

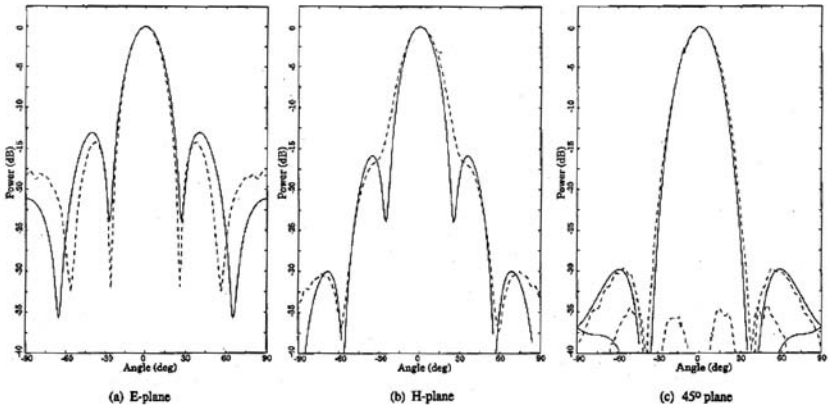


FIGURE 14-18 Measured and computed radiation patterns of a dielectric-loaded rectangular horn at 12.5 GHz with aperture geometry $a = 61$ mm, $b = 52.8$ mm, $d = 4.1$ mm, $\epsilon_{r1} = 3.07$, and $\epsilon_{r2} = 1$ (solid line — theory, dashed line --- experiment)

TABLE 14-4 Measured and Computed Gain of Dielectric-loaded Rectangular Horn with Aperture Geometry $a = 61$ mm, $b = 52.8$ mm, $d = 4.1$ mm, $\epsilon_{r1} = 3.07$, and $\epsilon_{r2} = 1$

Horn Type	12.25 GHz		12.5 GHz		12.75 GHz	
	Exp. dBi	Theory dBi	Exp. dBi	Theory dBi	Exp. dBi	Theory dBi
Unloaded	17.34	17.48	17.69	17.65	17.75	17.80
Dielectric-loaded	17.89	18.05	18.12	18.25	18.30	18.42

(experimental uncertainty ± 0.14 dB)

summarized in Table 14-4 for loaded and unloaded horns. The measured aperture efficiency in the loaded case is in excess of 91% across the band. The mismatch in a horn due to the dielectric loading is relatively small ($\epsilon_{r2} = 1$), although the bandwidth decreases and the aperture reflection increases as ϵ_{r1} is increased. However, for $\epsilon_{r1} < 4$ the effect is relatively small on the return loss. To obtain best results the dielectric loading was flared from the throat region to the aperture. Reduced efficiencies and lower sidelobes in both planes result when the dielectric loading on the wall of the transition is made too short.

14.3 CIRCULAR HORNS

Radiation from hollow circular pipes was investigated initially in the late 19th century by O. Lodge and J. C. Bose, and this was followed by theoretical work on circular waveguides by Lord Rayleigh.¹ However, it was not until the 1930s and 1940s that the first detailed studies of conical horns were undertaken, mainly by Southworth and King,^{1,28-29} the basis of current day design methodology. Because the circular waveguide is a cornerstone to this methodology, we summarize its radiation properties.

Radiation from Circular Waveguide

The radiation mechanism for a circular waveguide is identical to that for radiation from a rectangular waveguide, except that there is a different infinite set of TE and TM modes. The dominant mode of a circular waveguide is the TE_{11} mode, which has different polarization properties than the dominant mode of a rectangular waveguide—due mainly to the fact that the boundary conditions on the circular pipe cause the field lines to curve instead of be unidirectional. The radiated far-fields of the TE_{11} mode are given by

$$E_{\theta}(r, \theta, \phi) = 2\pi a \frac{e^{-jkr}}{r} J_1(k_c a) \frac{J_1(wa)}{k_c wa} \cos \phi \quad (14-44a)$$

$$E_{\phi}(r, \theta, \phi) = 2\pi a \frac{e^{-jkr}}{r} J_1(k_c a) \frac{k_c J_1'(wa)}{k_c^2 - w^2} \cos \theta \sin \phi \quad (14-44b)$$

where a is the aperture radius, $w = k \sin \theta$, and $k_c a = 1.841181$. The function J_n is the Bessel function of order n .¹²⁵ A prime on this function indicates the first derivative with respect to the argument. A more general form of Eq. 14-44 is available for circular waveguides with infinitely thin walls.¹¹⁰ This exact solution due to Weinstein¹¹⁰ finds wide application for feeds and measurement probes. The maximum gain of a circular waveguide operating in the TE_{11} mode only is

$$G_{\max} = 0.209(ka)^2 \frac{k}{\beta} \left(1 + \frac{\beta}{k}\right)^2 \quad (14-45)$$

where $\beta = \sqrt{k^2 - k_c^2}$ is the propagation constant. At frequencies well above cutoff ($\beta \approx k$), the input reflection coefficient Γ_{in} is small and $G_{\max} \approx 0.837(ka)^2$, giving an aperture efficiency of 83.7%. Choke rings can be used either inside the horn or at the aperture to tailor the radiation pattern of small horns to minimize cross-polarization over a moderate bandwidth.¹⁶ The E -plane beamwidth is wider than the H -plane up to a particular ka value. A choke ring placed concentric with the aperture with depth 0.2 to 0.26 λ can reduce the E -plane beamwidth to almost the same as the H -plane beamwidth.⁷¹ This method becomes less effective for larger diameter conical horns as the radiated field becomes too small to excite the ring slot. As a consequence, recourse must be made to other methods for achieving axisymmetric patterns. Some of these methods are described in the following sections.

Conical Horns

The conical horn excited by a circular waveguide in the TE_{11} mode is the counterpart of a pyramidal horn with TE_{10} excitation in a rectangular guide, and in many respects, its behavior is quite similar. The modes in an infinite conical horn are exactly expressible in terms of a combination of spherical Hankel and Legendre functions. Numerical methods such as mode matching, finite elements, or finite-difference time domain are usually preferred because accurate design information can be obtained for conical horns. Nevertheless, considerable physical insight is possible from a dominant mode approximation that parallels the rectangular horn description.

The geometry of the conical horn is given in Figure 14-19. Assuming the usual quadratic phase approximation, the fields radiated by this horn operating in the dominant mode is

$$E_\theta(r, \theta, \phi) = \frac{jE_o ka^2}{2} \frac{e^{-jkr}}{r} (G_0(w) - G_2(w)) \cos \phi \quad (14-46a)$$

$$E_\phi(r, \theta, \phi) = -\frac{jE_o ka^2}{2} \frac{e^{-jkr}}{r} \cos \theta (G_0(w) + G_2(w)) \sin \phi \quad (14-46b)$$

where $G_m(w) = 1/a^2 \int_0^a J_m(p_{11} \rho'/a) J_m(w\rho') \exp(-jk\rho'^2/2L) \rho' d\rho'$, $p_{11} = 1.841184$, and $w = k \sin \theta$. The above equations, obtained by Narasimhan and Rao,³¹ give useful results for semicone angles less than about 35° for medium-to-long horns ($L > 5\lambda$). For wider flare angles, the argument of the first Bessel function should be $(p_{11}/\theta_o a \tan(\rho'/L))$.

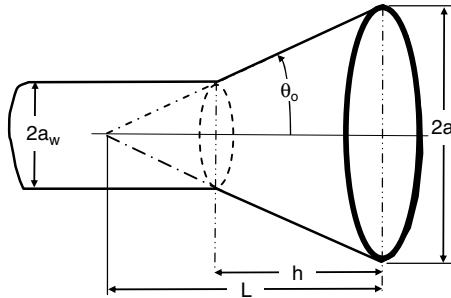


FIGURE 14-19 Geometry of conical horn

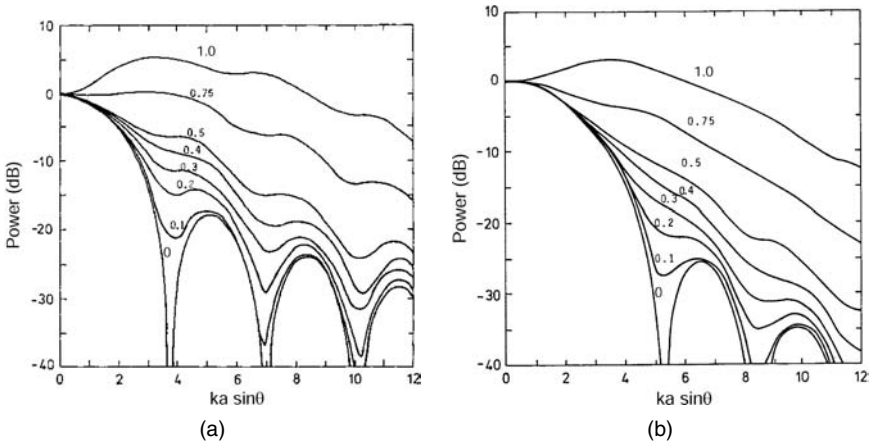


FIGURE 14-20 The universal radiation patterns of a conical horn in the (a) E -plane and (b) H -plane. The parameter is the normalized phase error Δ/λ .

The appropriate obliquity factors must be applied to Eq. 14-46 to suit the conditions at the aperture. The functions G_m are best evaluated by numerical integration, although a series solution is possible in terms of Lommel functions. As with the pyramidal horn, the radiation pattern of the conical horn is strongly dependent on the flare geometry. The E - and H -plane patterns are plotted in Figure 14-20 as a function of $\Delta = a^2/2L$, the distance in wavelengths between the spherical wavefront and the aperture. The quadratic phase increases as the semi-angle is increased, causing the patterns to broaden and the sidelobes to fill in. Figure 14-21 shows computed results and measured data from

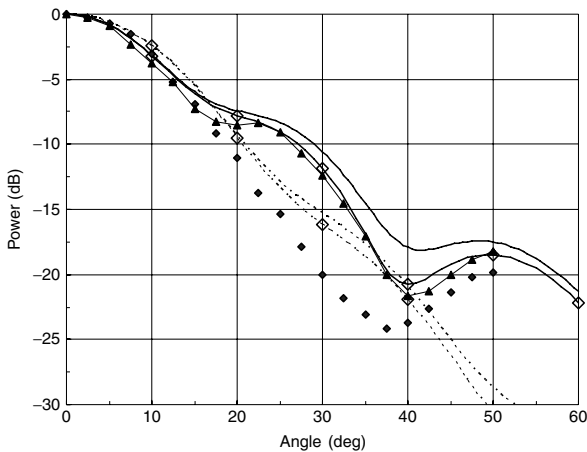


FIGURE 14-21 Principal plane patterns of a conical horn with $a = 1.7\lambda$, $L = 3.5\lambda$, and $\theta_0 = 25.91^\circ$: solid curve, E -plane; broken curve, H -plane; no marker, Eq. 14-46; open diamond, CST Microwave Studio; solid triangle, E -plane experiment (after King²⁹ © IEEE 1950); solid diamond, H -plane experiment (after King²⁹ © IEEE 1950).

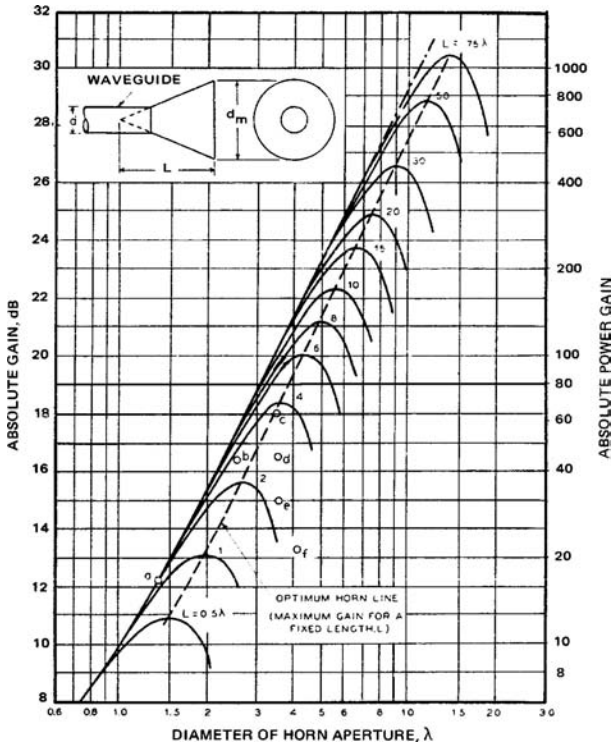


FIGURE 14-22 Calculated gain of a conical horn as a function of aperture diameter with axial length as parameter (after King²⁹ © IEEE 1950)

King for a horn with $a = 1.7\lambda$ and $L = 3.5\lambda$. Two sets of computed results are plotted: those given by Eq. 14-46 and predictions by CST Microwave Studio,¹²⁴ an accurate computer package. There is good agreement in the E -plane, although there is surprising discrepancy with King's measurements²⁹ in the H -plane. The computed return loss from Microwave Studio is >25 dB. The phase center in the E - and H -planes (given by Eq. 14-3) and the phase of Eq. 14-46 are plotted in Figure 14-22 as a function of Δ/λ between the -12 dB points of the main beam.

The absolute gain of a dominant-mode conical horn is

$$G_{\max} = \frac{4}{\beta/k} \frac{(ka)^2 |G_0(0)|^2}{J_1^2(p_{11}) (1 - 1/p_{11}^2)} \quad (14-47)$$

Expressed in dBi, this is

$$G(\text{dB}) = 20 \log(2\pi a/\lambda) - R_c \quad (14-48)$$

where R_c is analogous to the reduction factor for the sectoral horns described in Section 14.2, "Optimum-Gain Horns," that account for loss in gain due to spherical-wave phase error. The reduction factor is²

$$R_c \approx 0.8 - 1.71(\Delta/\lambda) + 26.25(\Delta/\lambda)^2 - 17.79(\Delta/\lambda)^3 \quad (14-49)$$

TABLE 14-5 Measured and Calculated Gain of Conical Horns

$L (\lambda)$	$a (\lambda)$	Calculated dBi	Expt. (King ²⁹) dBi
8.0	0.7	12.5	12.2
3.5	1.3	16.8	16.5
3.9	1.8	18.0	18.0
3.1	1.8	16.5	16.5
2.8	2.05	12.8	13.2
3.5	1.7	17.5	17.7

where $\Delta = a \tan(\theta_0/2)$. The computed (from Eq. 14-47) and measured gains²⁹ are listed in Table 14-5 for a variety of horn geometries and there is good agreement. The computed gain from CST Microwave Studio for the last horn (i.e., $L = 3.5\lambda$ and $a = 1.7\lambda$) is 17.73 dBi, which is in excellent agreement with King's measured result.

The design of a conical horn for specified beamwidths in the principal planes is performed iteratively. Assume an input waveguide of radius a_w , supporting only the TE_{11} mode. A design procedure is suggested here:

1. Starting with the patterns in Figure 14-20, select a trial value for the radius a to give an approximation to the desired half-beamwidth θ_{bw} in each principal plane. Select a value of horn axial length L , and calculate the length from $h = L(1 - a_w/a)$.
2. Compute the principal plane patterns using Eq. 14-46. Refine a and h to achieve the desired beamwidth.
3. Refine the design further (i.e., input match, polarization, etc.) by simulating the structure with an accurate horn software package.

It happens that for a small-diameter horn in which $a \approx 0.43\lambda$, the E - and H -plane patterns become nearly identical, at least down to the -12 -dB level, providing the wall thickness at the aperture is less than about $\lambda/100$. When the horn aperture terminates in a ground-plane, the diameter is larger for near identical patterns, when $a \approx 0.55\lambda$. In both cases, the half-beamwidth at -12 dB is about 65° and the peak cross-polarization in the intercardinal plane is < -30 dB, which makes the horn suitable as a feed for a reflector having $F/D = 0.4$.

Optimum-Gain Conical Horns

Most of the patterns shown in Figure 14-20 are for horns that are low gain as a result of significant aperture phase error. As mentioned in connection with pyramidal horns, an optimum horn refers to a horn whose aperture radius yields maximum gain for a given length. By means of Eq. 14-46, the maximum gain for a given length can be calculated for a conical horn. To a good approximation $a \approx \sqrt{0.7812 L \lambda}$, which corresponds to $\Delta = 0.39\lambda$, and the maximum gain expressed in dB is $G_{opt}(dB) \approx 20.29 \log(a/\lambda) + 12.85$. This means for an optimum conical horn the gain-reduction factor ranges from 2.5 dB to 3.2 dB as the axial length increases from short to long. This range corresponds to aperture efficiencies from 55.7% to 48.2%. Figure 14-22 shows the gain dependence on radius a in greater detail. These curves are historically interesting as King²⁹ gave them in his now classic work on conical horns. The plots differ slightly from the results of the expressions in Eq. 14-46 due to the employment of different methods and approximations in both instances.

Input Match

The input mismatch for conical horn Γ_{in} has similar properties to the pyramidal horn as described in the “Input Match” section within Section 14.2. The mismatch of dominant mode conical horns can be calculated from Eq. 14-30 with the admittance of the TE_{11} mode, given by

$$y_{11} = c_1 \frac{k}{\beta} \int_0^\infty \left(\frac{w\sqrt{1-w^2}}{(ka)^2} \frac{J_1'^2(kaw)}{(w^2 - (p_{11}/ka)^2)^2} + c_2 \frac{J_1^2(kaw)}{w\sqrt{1-w^2}} \right) dw \quad (14-50)$$

where $c_1 = 2p_{11}^4 / (p_{11}^2 - 1)$, $c_2 = 1 / p_{11}^4$, and $p_{11} = 1.841184$. Plots obtained from Eq. 14-50 show that Γ_{ap} is minimum when $ka \sim 3.5$ ($a \sim 0.56\lambda$), which is coincident with the condition for minimum cross-polarization for this mode.⁵⁵ Providing $a > \lambda$, $\Gamma_{ap} < -30$ dB. Design equations for the mismatch at the junction between a circular waveguide feeder and the flare section have been found unreliable. Therefore, it is recommended that once a preliminary design has been made, the horn should be analyzed using an accurate software package to obtain details of other horn characteristics such as mismatch. Suffice to say, the mismatch at the junction is low (< -20 dB) if the flare angle is relatively small ($\theta_o < 10^\circ$). The remedies described in Section 14.2, “Input Match,” can be used, if required, to reduce the mismatch at the junction and aperture.

Other Smooth-wall Circular Horns

Circular geometry has been rich in development of new structures and geometries. Some of these other types of circular horns are mentioned here. Circular corrugated horns are described in Section 14.4.

Coaxial Horns The coaxial waveguide and its horn counterpart are advantageous in some situations because the inside of the inner conductor can be used to support a waveguide or horn for high frequency operation (see Figure 14-23). An example is shown in Figure 14-1e. Such a method of dual-banding is limited to frequencies above the cutoff of the internal conductor. Other bands can be added by inserting additional coaxial waveguides inside successively smaller inner conductors, although it becomes progressively more difficult to feed each horn over a reasonable bandwidth. Letting the inner conductor radius be b and the outer radius be a , the frequency range of the inner conductor is limited to about 10% above the cutoff and bounded at the upper frequency by the TM_{11} mode (i.e., $-2 < kb < 3.83$). Another property of coaxial horns is that the operating bandwidth is slightly higher than a circular horn as the cutoff frequency of the dominant mode is less than that of circular waveguide. This cutoff is approximately given by $k_c a = 2\pi f_c a/c \approx 2/(1 - \tau)$, where f_c is the cutoff frequency and $\tau = b/a$. The bandwidth for TE_{11} mode operation is $2/(1 - \tau) < ka < \pi/(1 + \tau)$.

The main advantage of the coaxial horn is that the parameter τ provides an additional way of varying the radiation pattern, without incurring some phase loss due to the flare angle. Disadvantages of coaxial horns are the high sidelobes and significant aperture mismatch as τ is increased. Above about $\tau = 0.2$, the mismatch can be reduced using a pair of irises; the first one back from the aperture is connected to the outer conductor, and a second iris is connected to the inner conductor

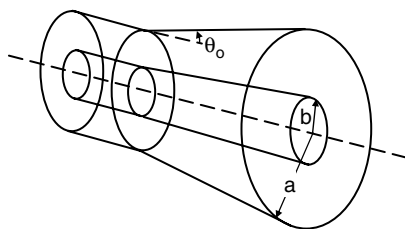


FIGURE 14-23 Geometry of coaxial horn

a farther distance back.⁵⁴ The higher sidelobes and mismatch are due to the annular aperture and the currents on the inner conductor. An approximate radiation pattern for conical coaxial horn is obtained by replacing the Bessel functions in Eq. 14-46 with the function

$$Z_m(x, x\tau) = J_m(x) - \frac{J'_m(x\tau)}{Y'_m(x\tau)} Y_m(x) \quad (14-51)$$

where x refers to the argument of the original Bessel function $J_m(x)$. In addition, p_{11} is now replaced by \bar{p}_{11} , the first root of $J_m(\bar{p}_{11})Y'_m(\bar{p}_{11}\tau) - J'_m(\bar{p}_{11}\tau)Y_m(\bar{p}_{11}) = 0$. These replacements also apply to the aperture admittance of the TE_{11} mode given by Eq. 14-50, where in addition the derivative of Eq. 14-51 (taken with respect to the first argument of $Z_m(x, x\tau)$ only) is also required. The maximum gain is computed from Eq. 14-47, wherein the denominator is replaced by $Z_1^2(\bar{p}_{11}, \bar{p}_{11}\tau)(1 - 1/\bar{p}_{11}^2) - \tau^2 Z_1^2(\bar{p}_{11}\tau, \bar{p}_{11}\tau)(1 - 1/(\bar{p}_{11}\tau)^2)$.

The H -plane beamwidth is approximately the same as for a conical horn, while the E -plane beamwidth increases with τ . The cross-polarization properties are similar to a circular waveguide except that the peak in the intercardinal plane increases, while the ka for minimum cross-polarization decreases with increasing τ . For a coaxial horn with a large flange, this condition is approximately given by

$$ka \approx 3.52 - 1.13\tau - 8.81\tau^2 + 16.34\tau^3 - 8.62\tau^4; \quad \tau \leq 0.8 \quad (14-52)$$

Multimode, Stepped, and Profiled Circular Horns We have seen that for rectangular horns the presence of the higher-order TM_{11} mode in addition to the dominant TE_{11} mode in a conical horn can create a radiation pattern with good axial-beam symmetry and with a low level of cross-polarization as well.¹⁷ When properly phased at the horn aperture, the TM_{11} mode cancels the ϕ -component of magnetic field due to the TE_{11} mode at the aperture boundary, causing H_ϕ as well as E_ϕ to vanish. This is a condition that leads to axial symmetry and low cross-polarization in the radiation pattern. In the original version (called the Potter horn⁷⁷), the TM_{11} mode was generated by a step change in guide radius from r_1 to r_2 in the horn throat, as shown in Figure 14-24. Of course, the radius r_1 must be large enough to support the TE_{11} mode but small enough to ensure that the TM_{11} is cut off (i.e., $1.84 < kr_1 < 3.83$). The amount of TM_{11} generated in this way increases with the ratio r_2/r_1 , but r_2 should not be so large as to permit the TE_{12} mode to propagate ($3.83 < kr_2 < 5.33$). Owing to the axial symmetry, TE_{mn} or TM_{mn} modes with $m > 1$ will not be excited. Figure 14-24 also indicates qualitatively the distribution of the electric field in the aperture for the two modes both individually and in combination. In the central region of the aperture the electric field of the TM_{11} mode reinforces that of the TE_{11} mode. Near the aperture boundary the two fields oppose one another. Thus the resulting electric field may be heavily tapered in the E - and H -planes.

A set of patterns in the E -, H -, and 45-degree planes is shown in Figure 14-25 for a horn with aperture diameter $2a = 4.67\lambda$ and a small semi-flare angle of 6.25° . In this horn $r_1 = 1.02\lambda$, $r_2 = 1.30\lambda$, and $\ell = 0.20\lambda$. The short constant-diameter section of length ℓ is used to ensure that the TE_{11} and TM_{11} modes will be in the same phase at the center of the aperture at the design frequency. This length depends upon the flare angle and the distance from the step to the aperture as well as on the launch phase of the TM_{11} mode relative to the TE_{11} mode.

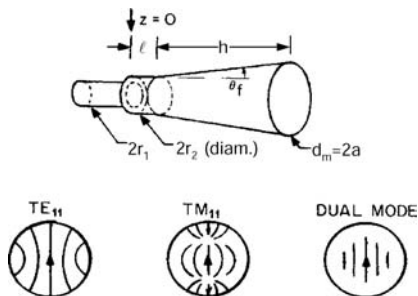


FIGURE 14-24 Dual-mode horn with generating step and its approximate aperture field distribution

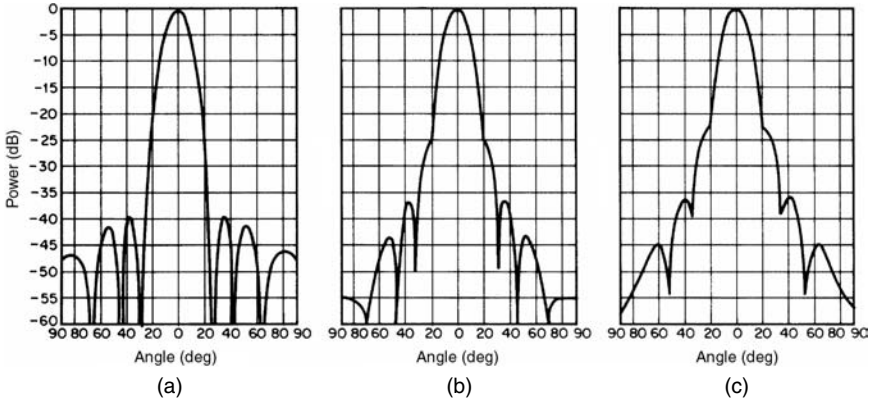


FIGURE 14-25 Radiation patterns of a dual-mode conical horn, $2a = 4.67\lambda$: (a) E -plane, (b) 45-degree copolar, and (c) H -plane (after Potter⁷⁷)

The TE_{11} mode generates θ - and ϕ -components of \mathbf{E} in the far-field, while the TM_{11} mode generates only the θ -component. In terms of the variable $w = ka \sin \theta$ and with the assumption that the aperture is reasonably large ($a \gg \lambda$) so that the reflection coefficient $\Gamma_{in} \approx 0$, the radiated far-field components are

$$E_{\theta} = \left[1 + \frac{\beta_{te}}{k} \cos \theta - \alpha \frac{\beta_{tm}/k + \cos \theta}{1 - (3.83/w)^2} \right] \frac{J_1(w)}{w} \cdot \cos \phi \quad (14-53a)$$

$$E_{\phi} = \left(\frac{\beta_{te}}{k} + \cos \theta \right) \frac{J'_1(w)}{1 - (w/1.84)^2} \cdot \sin \phi \quad (14-53b)$$

where $\beta_{te} = k\sqrt{1 - \left(\frac{1.84}{ka}\right)^2}$ and $\beta_{tm} = k\sqrt{1 - \left(\frac{3.83}{ka}\right)^2}$ refer to the propagation constants of the TE_{11} and TM_{11} modes respectively, and α , the mode-content factor, is a measure of the field strengths of the two modes at the aperture center. For $\alpha = 0.65$ the E - and H -plane HPBW's are equalized, and the phase centers in the planes are coincident. Other values of α , in the neighborhood of 0.6, lead to suppression of the first sidelobe in the E -plane or to suppression of the cross-polar radiation in the intercardinal planes. The presence of the TM_{11} mode has no effect in the H -plane, where this mode does not radiate. The beam-width of a dual-mode conical horn will vary somewhat with the mode-content factor; typically $HPBW \approx 36\lambda/a$.

Mode conversion at a discontinuous step in a round waveguide has been investigated theoretically and experimentally.¹²⁶⁻¹²⁷ Figure 14-26 shows the mode conversion coefficient and excitation phase of the TM_{11} mode relative to the TE_{11} mode on the guide center-line in the plane of the step at $z = 0$. If ϕ_{ex} represents the launch phase, ϕ_{ph} represents the differential phase in the constant-phasing section, and ϕ_{fl} represents the differential phase of the modes in the flared section of Figure 14-24, then the condition for reinforcement of the electric fields at the aperture center is

$$\phi_{ph} + \phi_{fl} - \phi_{ex} = 2m\pi, \quad m = 0, 1, 2, \dots \quad (14-54)$$

where

$$\phi_{ph} = \ell(\beta_{te} - \beta_{tm}) \quad (14-55)$$

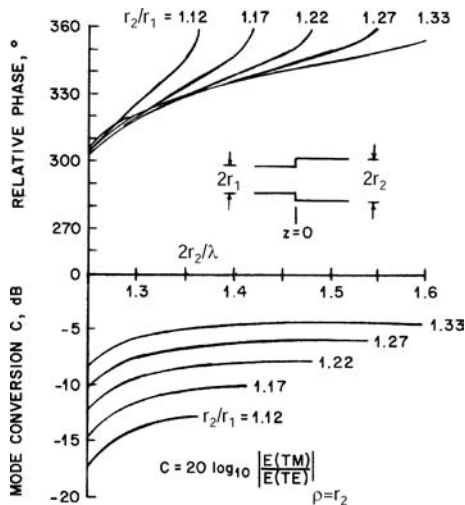


FIGURE 14-26 Mode conversion coefficient C and launch phase of TM_{11} mode relative to TE_{11} at step discontinuity (TM_{11} lags TE_{11}) (after Agarwal and Nagelberg¹²⁶ © IEEE 1970)

where β_{te} and β_{tm} depend on the radius r_1 of the phasing section. In the flared horn, the differential phase shift is closely given by

$$\phi_{fl} = \int_{z=\ell}^{z=\ell+h} (\beta_{te}(z) - \beta_{tm}(z)) dz \quad (14-56)$$

in which the guide propagation constants are now functions of the axial coordinate z and h , the length of the flare section. Note that ϕ_{ph} and ϕ_{fl} are positive angles because $\beta_{te} > \beta_{tm}$. These equations have been computed by Potter⁷⁷ and are reproduced in Figure 14-27. In the case of $a/\lambda = 2.33$, the z coordinate has been eliminated by introducing the semi-flare angle θ_f .

As an example of the use of these design curves, consider the horn whose patterns are given in Figure 14-25. With $a/\lambda = 2.33$, $r_2/\lambda = 0.65$, and $\theta_f = 6.25^\circ$, Figure 14-27a gives $\phi_{fl} \approx 625^\circ$, while a phasing section (r_2/λ) of length $\ell/\lambda = 0.25$ from Figure 14-27b gives $\phi_{ph} \approx 50^\circ$. The expected excitation phase is obtained from Figure 14-26; $\phi_{ex} \approx -45^\circ$ for $r_2/r_1 \approx 1.27$. When these values are substituted into Eq. 14-54, the equation is satisfied with a small error. From the same figure, the TM_{11} -mode conversion coefficient appears to be about -7.5 dB; evidently this corresponds closely to the value $\alpha = 0.6$ for the mode-content factor defined in Potter.⁷⁷

The phasing condition of Eq. 14-54 can be satisfied only at a single frequency because of the different dispersion characteristics of the two modes. Thus the dual-mode horn is frequency-sensitive (the horn in Figure 14-25 has a useful bandwidth of only 3 to 4 percent), and the sensitivity increases with horn length, which is typical for such horns. For horns that are not too long, some improvement can be obtained by loading the step discontinuity with a dielectric ring.¹²⁸

Use of a step discontinuity near the horn throat is not the only way to convert TE_{11} to TM_{11} energy in a horn. A technique described by Satoh¹²⁸ employs a thin dielectric band

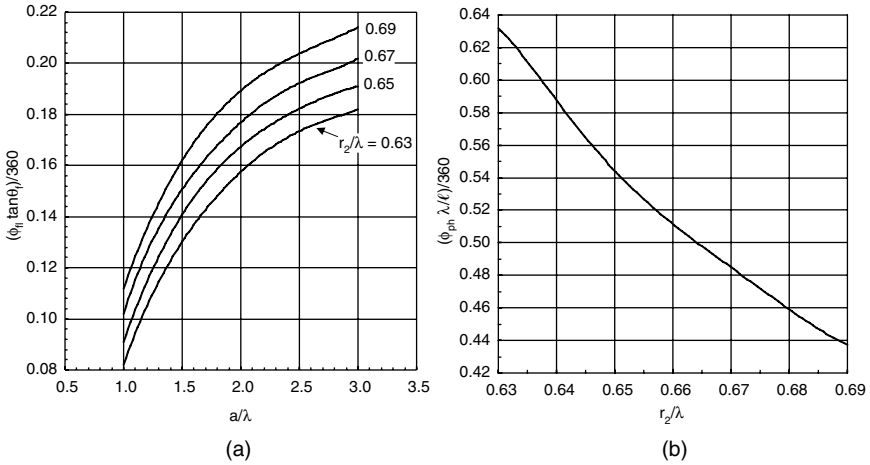


FIGURE 14-27 Differential phase shifts in (a) the flared horn and (b) the phasing section (after Potter⁷⁷)

placed internally in contact with the horn wall. When the length of the band, its thickness, and its position are correctly determined, the resulting bandwidth of this dual-mode horn is reported to be 25% (defined so that all sidelobes are less than -20 dB and E - and H -plane beamwidths differ by no more than 5%).

Conversion from the TE_{11} mode to the TM_{11} and other higher modes (TE_{12} , TM_{12} , TE_{13} , ...) at a relatively large-diameter junction between round and conical waveguides has been investigated analytically by Tomiyasu.¹²⁹ In this case, a discontinuous step (as in Figure 14-24) is not needed. Using this technique, Turrin¹³⁰ designed multimode horns of the type shown in Figure 14-28. The input guide is sized to propagate in the TE_{11} mode but not the TM_{11} mode ($1.84 < ka_w < 3.83$). Conversion takes place at the plane AA' in Figure 14-28, where the radius a is large enough to support this mode (i.e., $ka > 3.83$). The next-higher modes with the proper symmetry are the TE_{12} mode, which requires $ka > 5.33$, and the TM_{12} mode, requiring $ka > 7.02$. Since these modes are not wanted, the aperture diameter should be restricted so that $3.83 < ka < 5.33$ (i.e., $0.61 < a/\lambda < 0.85$).

At the plane AA' in Figure 14-28, the higher-order TM_{11} mode is 90° out of phase with the TE_{11} mode; hence, the phasing section ℓ should have a differential length of 270° for the two modes to ensure correct phasing at the aperture, thus

$$\ell(\beta_{te} - \beta_{tm}) = 3\pi/2 \quad (14-57)$$

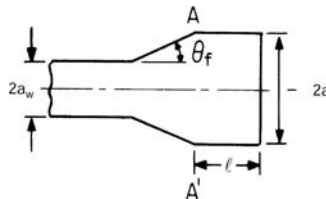


FIGURE 14-28 Stepless dual-mode horn (after Turrin¹³⁰ © IEEE 1967)

To ensure that the correct amount of TM_{11} relative to TE_{11} power is generated at the conical junction, the flare angle θ_f should be chosen to satisfy

$$\theta_f = 22.3\lambda/a \text{ degrees} \quad (14-58)$$

This simple equation comes from equating two different expressions for the mode power ratio, both given in Turrin.¹²⁹⁻¹³⁰ If Eq. 14-57 and Eq. 14-58 are applied to the two horn designs of Turrin,¹³⁰ they predict $\ell = 1.42\lambda$ and $\theta_f = 34^\circ$ for one horn, and $\ell = 3.86\lambda$ and $\theta_f = 24^\circ$ for another horn, in reasonably good agreement with the actual values.¹³⁰ Small horns constructed in accordance with these principles appear to yield a useful bandwidth of about 12%.

There is conversion of power from the TE_{11} mode to TM_{11} mode at the aperture of the horn in Figure 14-24 as well as at the step (also refer to “Multimode, Stepped, and Profiled Rectangular Horns”). The TM_{11} mode generated at the aperture is in quadrature with the TE_{11} component, and its amplitude is not negligible unless the flare angle is very small. The existence of two physically separate sources of TM_{11} -mode generation complicates the design of these horns, and recourse to empirical adjustment of step size and phasing length is usually necessary.

The methods described here can also be applied to conical coaxial horns. An example of the design and test of a Potter-type coaxial horn for a dual-band satellite application is described in James et al.⁵⁶ The difference compared with a circular conical horn is principally the higher sidelobe levels. The operating bandwidth in the lower frequency range is typically only a few percent.

The multimode circular horns described earlier are relatively narrowband. Wider-band horns can be designed by stepping the diameter upward in sections and flaring to the aperture using several axial corrugations. One example is shown in Figure 14-29. The aperture of this horn is relatively small and is suitable for feeding reflectors with $F/D \sim 0.4$. An edge illumination of -12 dB and lower is possible in both principal planes with good axial symmetry and, therefore, low cross-polarization over a 47% bandwidth (*input match* < -20 dB from 1.3 to 2.1 GHz). Through the selection of a number of internal axial corrugations (internal short-circuited ringslots) a combination of TE_{11} , TM_{11} , and TE_{12} modes is

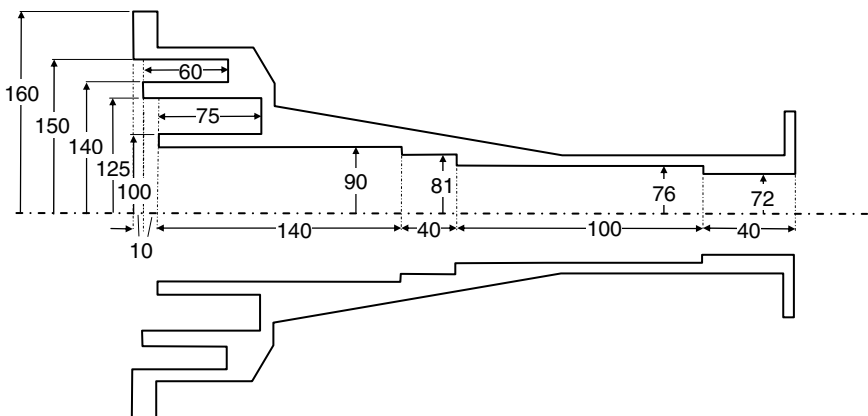


FIGURE 14-29 Stepped circular waveguide with axial corrugations for operation from 1.3 to 2.1 GHz. Dimensions are in mm.

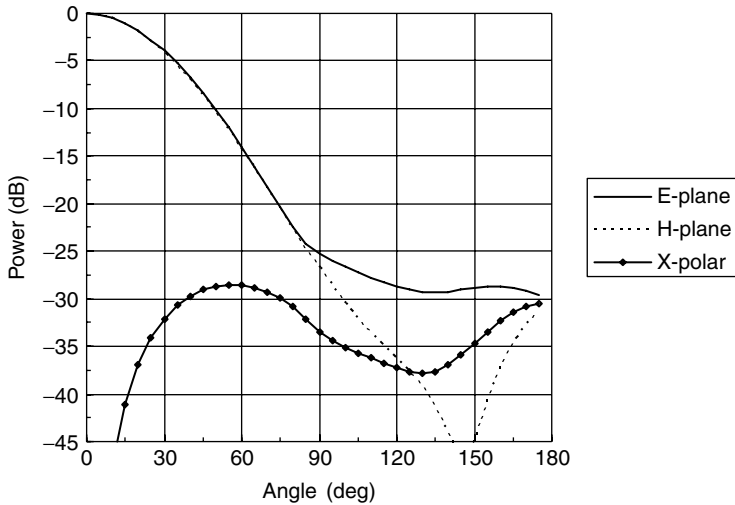


FIGURE 14-30 Computed radiation patterns at 1.75 GHz for the horn shown in Figure 14-29

produced in the aperture. These modes are supported by the TE_{11} mode in the coaxial corrugations. An example of the pattern produced in the intercardinal plane at the center of the band is shown in Figure 14-30, which has been computed using mode-matching software. At this frequency, the magnitude of the amplitudes of the TE_{11} , TM_{11} , and TE_{12} modes in the aperture are in proportion of 1:0.42:0.07 respectively. Toward the upper end of the band, the proportion becomes 1:0.54:0.19. The aperture radius is 150 mm, and the two axial corrugations are set back 10 mm from the aperture. The steps from the input to the axial corrugations are selected for the best input match from, in this case, a 72-mm input radius circular waveguide. The axial corrugations are chosen so that the innermost slot is 0.375λ deep at the center frequency, and the depth of the slots should reduce to about 0.25λ as the input of the shorted axial corrugations approaches the aperture. The end of the corrugations can be tapered linearly or have another shape (e.g., spherical). However, two axial corrugations are usually sufficient in many feed applications and little is gained with more than four. The geometry in Figure 14-29 can be scaled for operation in other frequency bands (i.e., scale the geometry by the factor $1.7/(\text{new center frequency in GHz})$).

Another similar horn is the Scrimp horn invented by Wolf⁸⁴ that is shown in Figure 14-31. This horn was originally developed for feed arrays onboard satellites. It has excellent aperture efficiency ($> 80\%$), an operating bandwidth of about 20%, low cross-polarization, and moderate sidelobe levels. There are three main parts to the Scrimp horn⁸⁴: a conical horn input section, a shorted coaxial waveguide, and a cylindrical output waveguide of length ℓ_c leading to the aperture. The input is a circular waveguide of diameter between 0.75λ to 0.95λ . When a TE_{11} mode is incident on the conical horn section, a standing wave is created in the coaxial waveguide section, which is dominated by a TE_{11} mode. The flare angle θ_c of the horn lies typically between 15° to 25° , and the length of the coaxial waveguide is $\ell_c \approx 0.25\lambda$. The TE_{11} mode in the conical horn oscillates in odd phase to the TE_{11} mode in the output section, and to ensure continuity of the fields additional higher-order modes, mainly TM_{11} and TE_{12} , are generated. These modes

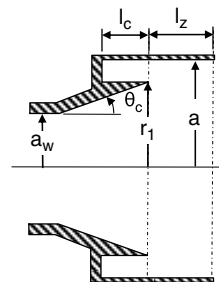


FIGURE 14-31 Geometry of the Scrimp horn

are superimposed on the fundamental TE_{11} mode to produce a field distribution in the aperture that is low in cross-polarized field components. The length ℓ_z is the projection of the flare to the aperture (i.e., $\ell_z \approx (a - r_i) \tan \theta_z$, although this length requires fine adjustment at each value of a to achieve the desired performance). The aperture diameter $2a$ can be between about 1.2 to 1.8λ depending on the desired aperture field taper.

An extension of the concept of tailoring the mode content in the aperture to synthesize a desired radiation pattern is to modify the profile of the horn between the input waveguide and output aperture. Several ways have been tried to control this process, although mostly they entail the use of an optimization method to carry out the changes necessary given a particular representation of the profile. The method of Deguchi et al⁶⁹ represents the horn as a sequence of cylindrical sections where the radius and length of each section is optimized to achieve the desired performance. Another approach, developed by Granet and Bird⁶⁸ and applied to smooth-wall horns by Granet et al,⁷⁰ represents the profile with cubic-splines, and the coefficients of the spline functions are optimized to meet specifications on the radiation pattern and input match. The profile is represented by only a few nodes to produce the spline, in particular two end nodes and five inner nodes. Only seven nodes are used to limit the number of optimization parameters to simplify the optimization process. Letting p_{jk} ($j = 1, \dots, N_p$) represent the power in the copolar pattern in direction j at frequency k ($k = 1, \dots, N_f$) and similarly q_{jk} is the power in the cross-polar pattern. Allow p_{ujk} and p_{ljk} to represent the upper and lower bounds on the copolar pattern and q_{djk} to specify the upper bound on the cross-polar level. Also, let Γ_k and Γ_{dk} denote the input reflection coefficient and the upper bound on reflection at frequency k . The aim of the optimization is to minimize a performance index based on the function

$$f_i = \begin{cases} \alpha_{ujk}(p_{jk} - p_{ujk}), i = 1, \dots, N_s; j = 1, 2, \dots, N_r; k = 1, 2, \dots, N_f \\ \alpha_{ljk}(p_{ljk} - p_{ljk}), i = N_s + 1, \dots, 2N_s; j = 1, 2, \dots, N_r; k = 1, 2, \dots, N_f \\ \alpha_{xjk}(q_{jk} - q_{djk}), i = 2N_s + 1, \dots, 3N_s; j = 1, 2, \dots, N_r; k = 1, 2, \dots, N_f \\ \alpha_{rk}(\Gamma_k - \Gamma_{dk}), i = 3N_s + 1, \dots, N_o; k = 1, 2, \dots, N_f \end{cases} \quad (14-59)$$

where $N_s = N_r N_f$, $N_o = 3N_s + N_f$, $i = 1, \dots, N_o$, α_{ujk} , α_{ljk} , etc., are weighting factors on the copolar, cross-polar, and reflection coefficient responses. There are many ways of seeking zero or negative values for f_i , such as the least p-th index and genetic algorithms. The least p-th index was used in conjunction with Fletcher's quasi-Newton search method by Granet and coworkers^{68,70} and found effective. By these means, circular horns with many different properties can be designed. For example, a horn with a gain of about 25 dBi has been designed to operate over a 50% bandwidth.⁷⁰ The horn profile is shown in Figure 14-32 and the measured and calculated results for a horn developed for the 100-GHz band is given in Figure 14-33.⁷⁰ The profile is close to a $\sin^{3/4}$ -profile followed by a Gaussian profile extension (shown as a dotted line in Figure 14-32). The performance of this horn compares favorably with the more complicated corrugated horn, which turns out to have a shorter length and the radiation patterns to have lower sidelobes. However, smooth-wall horns are simpler to manufacture, especially at high frequencies, and in some applications their performance may suffice.

As we have seen, the Gaussian⁶⁶ (also called *hyperbolic*⁶⁵) profile is very useful for completing the transition of the horn flare from the mid-range matching section to the aperture. Also, it results in compact horns compared with a linear taper. This profile is defined as a function of axial distance (z) and is given by

$$a(z) = a_i \sqrt{1 + \left(\frac{z}{L_p}\right)^2 \left(\left(\frac{a_o}{a_i}\right)^2 - 1 \right)} \quad (14-60)$$

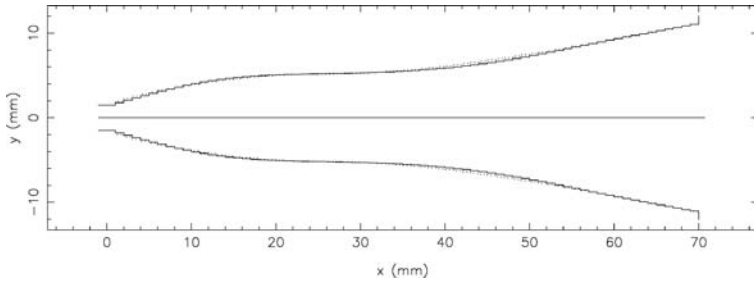


FIGURE 14-32 Profile of smooth-wall horn designed using splines and optimization. Dimensions are given mm.

where a_i is the radius of the input section, a_o is the aperture radius, and L_p is the length of the profiled section.

Dielectric-lined Conical Horns A conical horn with dielectric loading, as shown in Figure 14-34, is capable of operating over very wide bandwidths. A low-dielectric permittivity material or air, with permittivity ϵ_{r1} , between a core of permittivity ϵ_{r2} and a metallic wall allows the creation of a hybrid HE_{11} mode that has very attractive radiation properties, such as good axial symmetry, low sidelobes, and low cross-polarization. There is an optimum gap width t between the core and the wall, and this is given approximately by⁴⁷

$$t \approx \frac{\lambda}{4\pi\sqrt{\epsilon_{r2} - \epsilon_{r1}}} \ln \left(\frac{\sqrt{\epsilon_{r2}} + \sqrt{\epsilon_{r1}}}{\sqrt{\epsilon_{r2}} - \sqrt{\epsilon_{r1}}} \right) \quad (14-61)$$

The hybrid-mode of a conical horn has a low frequency cutoff determined mainly by the permittivity of the core, and the upper frequency is limited by the cutoff of the EH_{12} mode.⁴⁵ This means the performance can be extremely wideband, as much as in excess of 30:1,⁴⁹ although as one might expect, performance is not constant over the band and the beamwidth reduces with increasing frequency.

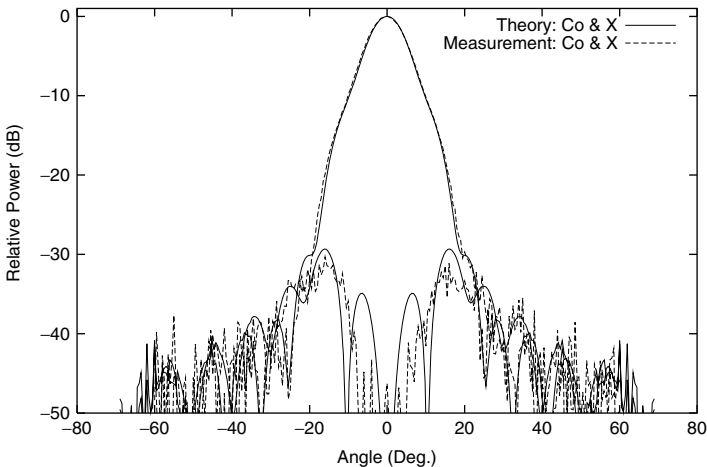


FIGURE 14-33 Radiation patterns in the intercardinal plane of the horn shown in Figure 14-32 at 107 GHz (after Granet et al⁷⁰ © IEEE 2004)

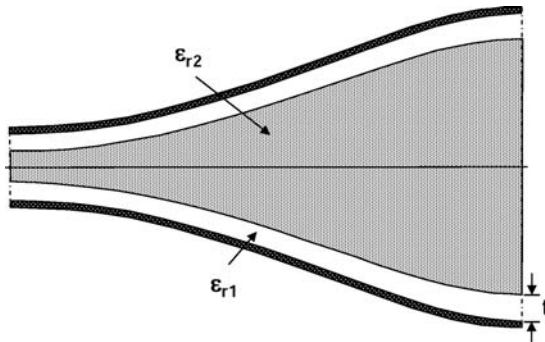


FIGURE 14-34 Geometry of dielectric-loaded conical horn

The literature describes many different types of dielectric-loaded circular horns.⁴⁵⁻⁵¹ The horn can have a solid-core dielectric with a permittivity of about 2.5 or foam materials with permittivity close to 1. However the widest bandwidth horns use a dielectric permittivity of between 1.1 and 1.2 for the inner core and also a thin-air gap given by Eq. 14-61.

Loss in a dielectric horn can be significant if the material properties are not carefully selected. A simulated permittivity can be made to achieve the desired permittivity by placing regular holes in a selected material such as PTFE, loading a polystyrene foam with small high permittivity spheres, or using alternating layers of low- and high-permittivity materials.⁵⁰ Care must be taken to ensure symmetry is not lost by this approach (equivalent to introducing anisotropy) or that the glue used to bind the materials is not contributing unnecessary loss.

14.4 CORRUGATED HORNS

For many years after the introduction of microwave frequencies, the smooth-wall circular and rectangular horns were the only known types. However, during the 1940s and 1950s the benefits of corrugating surfaces became apparent. Kay¹ was the first to describe a conical corrugated horn and shortly afterward, simultaneously research teams in the United States³⁴ and Australia³⁵ announced publicly the beneficial effects of transverse corrugations on the performance of circular horns. This led to detailed study of corrugated horns⁴¹ that continues to this day. This section summarizes the important information for the design of corrugated horns of circular and rectangular cross-section. We begin with a tutorial on corrugated surfaces.

Hybrid Modes: Soft and Hard Surfaces

A mode in a structure such as a waveguide or a horn that has non-zero field components in the direction of propagation is called *hybrid* to distinguish it from the usual TE or TM modes in hollow pipes. Natural structures, such as dielectric rods, support only hybrid modes. There are also a multitude of other structures that support hybrid modes that occur as a consequence of requiring axial-beam symmetry, low sidelobes, and low cross-polarization. These properties can be obtained with some structures over a wide frequency range (16:1 or more) in a conical horn or open-ended round waveguide that supports the hybrid HE_{11} mode. This mode is a mixture of TE_{11} and TM_{11} modes of the form $TM_{11} + \gamma TE_{11}$, where γ is called the mode-content factor. The waveguide wall presents an impedance to the fields, which is a pure

reactance in the lossless case. Surface impedances are defined for a direction parallel to the surface, say ζ , as follows:

$$Z_\zeta = \frac{\hat{\zeta} \cdot \mathbf{E}}{\mathbf{H} \cdot (\hat{n} \times \hat{\zeta})} \bigg|_{\text{surface}} \quad (14-62)$$

where \hat{n} is the normal to the surface and orthogonal to $\hat{\zeta}$. For the most desirable horn performance, the inner surface of the horn or guide must be anisotropic in such a way that it has different reactances, X_ϕ and X_z in the azimuthal and z -directions. When these conditions are realized, the TE and TM components become locked together as a single (hybrid) mode, and they propagate with a unique common velocity.

Such an anisotropic impedance can be achieved by cutting circumferential slots into the surface of a metallic waveguide. Thus, in Figure 14-35, which represents a longitudinal section through a cylindrical corrugated waveguide, E_ϕ will be zero on the surface of the teeth at $\rho = a$. Each slot acts as a radial transmission line that is short-circuited at $\rho = a + d$. If the line length $d = \lambda/4$, then an open circuit appears across each gap at $\rho = a$. Providing there are several slots per wavelength, then the conditions $X_\phi \rightarrow 0$ and $X_z \rightarrow \infty$ will be realized in an average sense. The open-circuit gaps ensure that no axial current can flow at $\rho = a$. Consequently, the azimuthal magnetic field component, which induces axial current flow, is zero. On average, the same boundary conditions now prevail for both azimuthal field components and, therefore, $E_\phi = 0$ and $H_\phi = 0$ on the guide walls. This condition on the aperture field of a horn leads to axial symmetry and low cross-polarization in its radiation pattern.

The hybrid mode in the waveguide has the cylindrical electric field components

$$e_\rho(\rho, \phi) = -j \left[\frac{\beta}{k} J_1'(k_\rho \rho) + \gamma \frac{J_1(k_\rho \rho)}{k_\rho \rho} \right] \cos \phi \quad (14-63a)$$

$$e_\phi(\rho, \phi) = j \left[\frac{\beta}{k} \frac{J_1(k_\rho \rho)}{k_\rho \rho} + \gamma J_1'(k_\rho \rho) \right] \sin \phi \quad (14-63b)$$

$$e_z(\rho, \phi) = \frac{k_\rho}{k} J_1(k_\rho \rho) \cos \phi \quad (14-63c)$$

where $\exp(-j\beta z)$ is understood, and β and k_ρ are the propagation constants in the z and ρ directions, respectively, and are related by $\beta^2 + k_\rho^2 = k^2$. The magnetic field components can be found from Maxwell's equations in the usual way. It is noted that the ratio of the

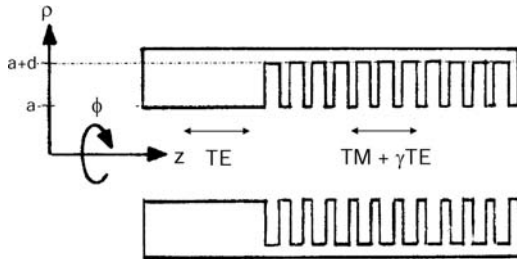


FIGURE 14-35 Corrugated circular waveguide with TE_{11} -mode excitation

axial field components is $e_z/h_z = \gamma$. Expressed in rectangular coordinates, the transverse electric field components are

$$e_x(\rho, \phi) = \frac{E_o}{2} [(1 + \gamma k/\beta) J_0(k_\rho \rho) - (1 - \gamma k/\beta) J_2(k_\rho \rho) \cdot \cos 2\phi] \quad (14-64a)$$

$$e_y(\rho, \phi) = -\frac{E_o}{2} (1 - \gamma k/\beta) J_2(k_\rho \rho) \sin 2\phi \quad (14-64b)$$

where $E_o = j\beta/k$. It is clear that the field will be polarized in the x -direction only (i.e., zero cross-polarization) if $\gamma = \beta/k$. Furthermore, this condition ensures that e_x tapers radially with ρ but does not vary with ϕ . However, if $\gamma = -\beta/k$, the transverse field is a mixture of e_x and e_y with strong azimuthal variations.

To satisfy the condition $X_\phi = 0$, $E_\phi = 0$ at $\rho = a$, and then Eq. 14-63 gives

$$\gamma = -\frac{\beta}{k} \cdot \frac{J_1(k_\rho a)}{k_\rho a J_1'(k_\rho a)} = \frac{\beta}{k} \cdot \frac{J_2(k_\rho a) + J_0(k_\rho a)}{J_2(k_\rho a) - J_0(k_\rho a)} \quad (14-65)$$

Thus, it is observed that the condition $\gamma = \beta/k$ occurs at the roots of $J_0(k_\rho a)$ (i.e., for $k_\rho a = 2.4048, 5.5201, 8.6537, \dots$ corresponding to the HE_{1n} ($n = 1, 2, 3, \dots$) modes, respectively). The alternative condition $\gamma = -\beta/k$ occurs at the roots of $J_2(k_\rho a)$ (i.e., for $k_\rho a = 0, 5.1356, 8.4172, \dots$, and these modes are designated EH_{1n} ($n = 0, 1, 2, \dots$), respectively).

The longitudinal reactance X_z (which is infinite at the slot resonant frequency) is obtained from Eq. 14-62 as $jX_z = e_z / h_\phi|_{\rho=a}$. Hence, from Eq. 14-63

$$X_z/\eta = \frac{-k_\rho^2 a J_1(k_\rho a)}{k k_\rho a J_1'(k_\rho a) + \gamma \beta J_1(k_\rho a)} \quad (14-66)$$

where η is the free-space wave impedance. Combining Eqs. 14-65 and 14-66 gives

$$\frac{1}{\gamma} - \gamma = \frac{k_\rho^2 a}{\beta} \cdot \eta/X_z \quad (14-67)$$

Consequently, in the special case of resonant slots ($X_z = \infty$), it follows that $\gamma = \pm 1$. This is called the *balanced hybrid condition*, and, therefore, $\gamma = 1$ for HE modes and $\gamma = -1$ for EH modes. Because Eq. 14-67 is a quadratic in γ , it has two roots in the general case and the designations HE and EH still apply even under unbalanced conditions. The characteristic equation for HE or EH hybrid modes is obtained from Eq. 14-66 by substituting for γ in Eq. 14-65. The surface reactance is capacitive over most of the preferred operating range of the corrugated horn from about where $X_z = \infty$ to $X_z = 0$.

The characteristic equation for the m -th order HE_{1n} mode⁴¹ is

$$\Lambda_n(k_\rho a) - \frac{\left(\frac{n\beta}{k}\right)^2}{\Lambda_n(k_\rho a)} = \left(\frac{k_\rho}{k}\right)^2 S_n(ka, k(a+d)) \quad (14-68)$$

where a is the inner radius of the corrugated waveguide, $a+d$ is the radius to the base of the slot (see Figure 14-35), $\Lambda_n(x) = x J_n(x)/J_n'(x)$, and a function related to the surface admittance presented by the slots at $\rho = a$ given by⁴¹

$$S_n(x, y) = x \frac{J_n'(x) Y_n(y) - J_n(y) Y_n'(x)}{J_n(x) Y_n(y) - J_n(y) Y_n(x)} \quad (14-69)$$

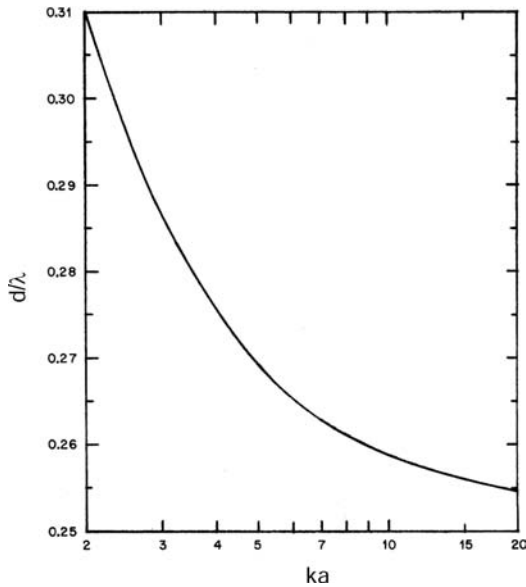


FIGURE 14-36 Slot depth d at resonance (after Thomas³⁹
© IEEE 1978)

Study of Eq. 14-68 shows that under the desired condition on the transverse field for the dominant mode ($n = 1$), namely $k_\rho a = 2.4048$, is approached asymptotically when the normalized frequency ka is large. In that case, β/k approaches unity, and from Eq. 14-65, γ must approach +1. Further, when ka is large, $k_\rho a$ will be close to 2.4048 even when X_z is far from resonance. This means that $J_0(k_\rho a)$ will be small and that $\gamma k/\beta$ will not depart greatly from unity despite the fact that X_z changes drastically with frequency. This means that the balanced hybrid condition is approximated over a wide range of frequencies of up to about 2.2:1. Operation is limited at the upper end of the band by the appearance of the unwanted EH_{11} mode.

Slot resonance ($X_z = \infty$) occurs when the slot depth d is one-quarter wavelength in the radial line of radii a and $a + d$. This depth $d \approx \lambda/4$ when ka is large, is shown in Figure 14-36. The tooth width t in Figure 14-37 should be small compared with the slot width w , typically $t < 0.2w$, in order to reduce frequency sensitivity. There should be several slots per wavelength, preferably

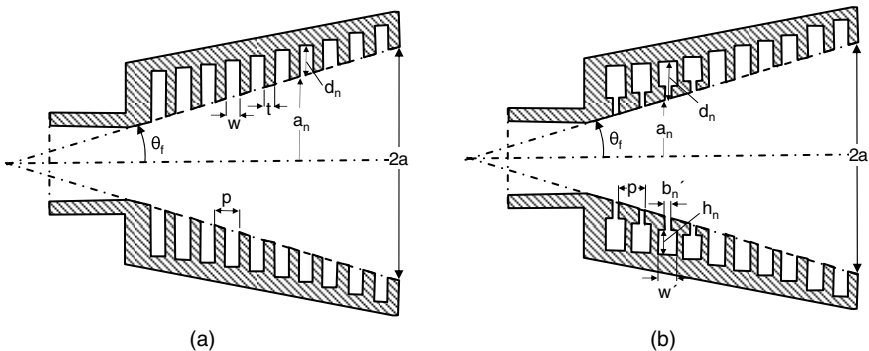


FIGURE 14-37 Small flare angle corrugated horn: (a) conventional corrugations and (b) ring-loaded slots at the input

four or more, although in wideband horns it is permissible to have only two per wavelength at the high-frequency end of the band.

In acoustics, surfaces are termed either hard or soft on the basis of whether there is intense power flow along the surface or no power flow at all close to the surface. There is an analogous situation for electromagnetic waves.¹³¹⁻¹³⁴ The inner surface of the corrugated horn described previously is an example of a soft surface at the slot resonant frequency. Hard surfaces also exist in electromagnetics. An example is a longitudinally corrugated surface in which the slots are filled with dielectric.

For the corrugated horn, the surface impedances are anisotropic; the field in the horn satisfies the general relationship for axial-pattern symmetry, that is

$$Z_\phi Z_z = -X_\phi X_z = \eta^2 \quad (14-70)$$

This is satisfied only if $Z_z = \infty$, and therefore $h_\phi = 0$ at the surface, as was pointed out previously.

Because both e_ϕ and h_ϕ are zero at the inner surface of the corrugated horn under balanced conditions, the axial component of the Poynting vector at this point must also be zero, and there will be no axial power flow along the surfaces. The corrugated surface is, therefore, soft, and the horn itself is sometimes referred to as a *soft surface horn* or simply a *soft horn*. As indicated previously, the surface reactance conditions for a soft horn are

$$X_\phi = 0 \quad \text{and} \quad X_z = \infty \quad (14-71)$$

By contrast, for a hard surface the conditions are the reverse (dual) of these, namely,

$$X_\phi = \infty \quad \text{and} \quad X_z = 0 \quad (14-72)$$

The above is the condition for a so-called *hard horn*. Further discussion of hard horns is delayed until Section 14.5, in the section "Horns with Hard Surfaces."

Conical Corrugated Horns

The radiation from a conical corrugated horn operating in the HE_{11} dominant mode near the balanced hybrid condition can be obtained from aperture fields given by Eq. 14-64 with a quadratic phase factor given by Eq. 14-5 under the usual assumptions ($\theta_0 < 35^\circ$). The result for the far-fields is

$$E_\theta(r, \theta, \phi) = \frac{jE_o ka}{2} \frac{e^{-jkr}}{r} F_\theta [(1 + \gamma k/\beta)G_0(w) + (1 - \gamma k/\beta)G_2(w)] \cos \phi \quad (14-73a)$$

$$E_\phi(r, \theta, \phi) = \frac{-jE_o ka}{2} \frac{e^{-jkr}}{r} F_\phi [(1 + \gamma k/\beta)G_0(w) - (1 - \gamma k/\beta)G_2(w)] \sin \phi \quad (14-73b)$$

where $F_{\theta,\phi}$ are the obliquity factors, G_m is given by the integral following Eq. 14-46 wherein the argument of the Bessel function p_{11} is replaced by $k_\rho a$, which is found in general from the dispersion relation (Eq. 14-68) for a specified k , aperture radius, and slot depth d (see Figure 14-36). The other parameters required are $\beta/k = \sqrt{1 - (k_\rho/k)^2}$ and γ , which follow from Eq. 14-65. However, under balanced hybrid conditions $\gamma k/\beta = 1$, and for the dominant mode, $k_\rho a = 2.4048$, allowing Eq. 14-73 to simplify to

$$E_\theta(r, \theta, \phi) = E_o ka \frac{e^{-jkr}}{r} F_\theta G_0(w) \cos \phi \quad (14-74a)$$

$$E_\phi(r, \theta, \phi) = -E_o ka \frac{e^{-jkr}}{r} F_\phi G_0(w) \sin \phi \quad (14-74b)$$

which for a directive beam imply cross-polarization is almost negligible and the copolar main beam is almost axially symmetric for the HE_{11} mode for angles θ close to the axis ($< 50^\circ$). However, the rapid decay of the G_0 function with θ usually means for most horns that the obliquity factors have only a secondary effect on both co- and cross-polarization. This is also true for the higher-order HE_{12} , HE_{13} , ... modes. The balanced EH modes, which correspond to $J_2(k_\rho a) = 0$ and $\gamma = -\beta/k$, radiate with a null in the axial direction ($\theta = 0$) and a state of polarization that may be linear, elliptical, or circular, depending on ϕ . They are almost invariably undesirable in antenna engineering applications.

Universal patterns for the small to medium flare horn under balanced hybrid conditions and large ka are plotted in Figure 14-38. The phase center in the principal planes as a function of $k\Delta$, where $\Delta = a^2/2L$, evaluated under the same conditions is, of course, identical, and the result is shown in Figure 14-6. The curve was calculated from Eq. 14-3 and the phase of Eq. 14-74 between the -12 dB points of the main beam. According to Thomas³⁹ when ka is less than 12, the sidelobe level for a given Δ gradually decreases but the shape of the main lobe is virtually unaffected until ka reaches about 6. For $ka < 6$, the flange at the aperture begins to affect the pattern. The patterns become invalid for $ka < 4$. When the flange affects the pattern shape (i.e., $ka < 6$), a remedy recommended in Thomas³⁹ is to extend the corrugations into the plane of the flange. This will preserve pattern symmetry, but it increases the effective aperture size slightly. The new aperture radius is approximately $a' = a + e/2$, where e is the radius of the curved portion of the flange.

The conical corrugated horn has certain advantages over the dominant-mode pyramidal horn as a gain standard. It is not subject to the small gain uncertainties caused by E -plane edge diffraction; its ohmic loss is very small; and it has considerably lower sidelobe levels than pyramidal and smooth-wall conical horns. The absolute gain of a small-to-medium flare dominant mode corrugated horn is

$$G_{\max} = \frac{4\beta/k (ka)^2 |(1 + \gamma k / \beta) G_0(0)|^2}{\left(1 + \frac{\gamma\beta}{k}\right) \left(1 + \frac{\gamma k}{\beta}\right) (J_0^2(k_\rho a) + J_1^2(k_\rho a)) + \left(1 - \frac{\gamma\beta}{k}\right) \left(1 - \frac{\gamma k}{\beta}\right) (J_2^2(k_\rho a) - J_1(k_\rho a)J_3(k_\rho a))} \quad (14-75)$$

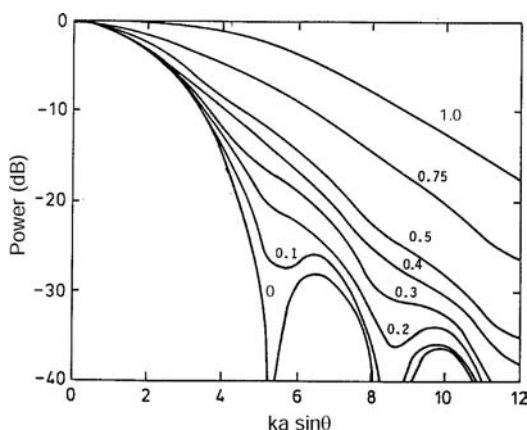


FIGURE 14-38 Universal patterns for small-flare-angle corrugated horns under near-balanced conditions. The parameter is the normalized phase error Δ/λ .

This simplifies considerably under balanced hybrid conditions to

$$G_{\max} = 4(ka)^2 \frac{|G_0(0)|^2}{J_1^2(k_p a)} \quad (14-76)$$

The maximum gain for small flare angles ($\theta_0 \sim 0$) with $k_p a = 2.4048$ is approximately $0.692(ka)^2$, that is an aperture efficiency of 69.2%. The gain (Eq. 14-75) expressed in dBi is

$$G(\text{dB}) = 20 \log(2\pi a/\lambda) - R_s \quad (14-77)$$

where R_s is the gain reduction factor, which is a function of $k_p a$. Some polynomial approximations of R_s as a function of the phase factor Δ at selected values of $k_p a$ are given here¹³⁵:

$$\begin{aligned} 1.3351 - 1.3341(\Delta/\lambda) + 14.874(\Delta/\lambda)^2 - 3.5629(\Delta/\lambda)^3; & k_p a = 2.3 \\ R_s \approx 1.6307 - 0.7893(\Delta/\lambda) + 11.667(\Delta/\lambda)^2 - 2.1607(\Delta/\lambda)^3; & k_p a = 2.405 \\ 2.0038 - 0.2956(\Delta/\lambda) + 8.2429(\Delta/\lambda)^2 - 0.3715(\Delta/\lambda)^3; & k_p a = 2.5 \end{aligned} \quad (14-78)$$

The gain is insensitive to deviations from the zero cross-polar condition, $k_p a = 2.4048$, and this is particularly true in the vicinity of $\Delta \approx 0.48\lambda$.

The radiation pattern of a typical conical corrugated horn is shown in Figure 14-39. The horn is pictured in Figure 14-1d. It has radius $a = 140$ mm, semi-angle = 12.5° , length $h = 560.0$ mm and is fed from a waveguide of radius $a_w = 11.16$ mm.¹³⁶ There are 100 variable depth corrugations with a pitch 6 mm and slot-to-pitch ratio of 0.75. It operates satisfactorily across the band from 10.5 to 14.5 GHz and has a gain of 25.8 dBi at 12.5 GHz. The measured cross-polar level in the band is < -35 dB, and the return loss is > 25 dB. Results computed from Eq. 14-73 for a slot depth of 6.5 mm and by a mode-matching method are shown in Figure 14-39 also.¹³⁶

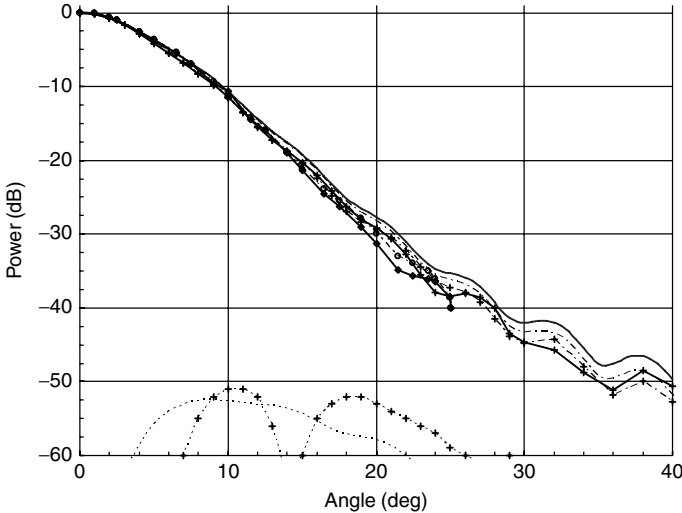


FIGURE 14-39 Radiation pattern of conical corrugated horn at 12.5 GHz. The horn has radius $a = 140$ mm, length $h = 560.0$ mm, semi-angle = 12.5° , and it is fed from a waveguide of radius $a_w = 11.16$ mm. The solid line represents the E -plane, the broken line the H -plane, and the dashed line is the cross-polar pattern in the 45° plane. No symbol represents the dominant mode model, + represents mode matching, and o the experimental results.

A useful property of corrugated horns is that the radiated field can be expressed as a small number of Gaussian modes. This fact is very helpful in the design of quasi-optical systems for millimeter and submillimeter-waves.¹³⁷ Gaussian modes have a Gaussian function amplitude variation with radial distance from the axis of propagation; the decay depends on the distance z along the beam from the source. The Gaussian distribution is the first term of a Laguerre-Gaussian expansion, which allows the radiated field of a corrugated horn to be expressed as a series of such modes.¹³⁸ The zero-order term of this series for the unidirectional field of the HE_{11} mode under balanced-hybrid conditions has the following Gaussian beam-mode propagation function:

$$e_x(\rho, z') = \frac{w_o}{w(z')} \exp \left[-jk \left(z' + \frac{\rho^2}{2R_c(z')} \right) - \frac{j\rho^2}{w(z')^2} + j\phi_c(z') \right] \quad (14-79)$$

where z' is the distance along the beam, ρ is the distance from the beam axis, $w(z') = w_0 \sqrt{1 + \left(\frac{z'}{z_c} \right)^2}$ is the beam radius, $R_c(z') = \left(z' + \frac{z_c^2}{z'} \right)$ is the radius of curvature of the wavefront, $z_c = \frac{\pi w_0^2}{\lambda}$ is the confocal distance, w_0 is the beam waist radius and $\phi_c(z') = a \tan(z'/z_c)$ is the phase shift. The HE_{11} mode has been normalized for unit power. Let $z' = z + z_a$ where z is the aperture coordinate and z_a is the distance of the Gaussian mode origin to the aperture. It can be shown¹³⁷ that for maximum coupling from the HE_{11} mode to the zero-order Gaussian mode $w(z_a) = 0.6435a$. Also, the radius of curvature $R_c(z_a) = L$. These determinations allow the Gaussian beam for the corrugated horn to be fully specified since now

$$w_0 = 0.6435a / \sqrt{1 + \left(\frac{0.4141\pi a^2}{\lambda L} \right)^2} \quad \text{and} \quad z_a = L / \left[1 + \left(\frac{\lambda L}{0.4141\pi a^2} \right)^2 \right] \quad (14-80)$$

Other horns can be represented in terms of Gaussian modes, although many modes are needed to give an accurate approximation of the beam, which complicates the design. A detailed description of Gaussian beam-mode techniques and applications can be found in Goldsmith.¹³⁷

Optimum-gain Corrugated Horns

In common with the other horns, there is a family of optimum corrugated horns. These horns provide the maximum gain for a given length. From Eq 14-75 and with $k_\rho a = 2.4048$, to a good approximation $a \approx \sqrt{0.976 L \lambda}$, and the maximum gain in dB is $G_{\text{opt}}(\text{dB}) \approx 20 \log(a/\lambda) + 12.06$. The phase factor is $\Delta \approx 0.48\lambda$, which means that for an optimum horn the aperture efficiency is about 40.7%. Near the optimum region, the gain is insensitive both to frequency and to small deviations from the condition $k_\rho a = 2.4048$.

Wide-Flare-Angle Scalar Horn

A wide-flare horn is shown in Figure 14-40, and it is apparent that the path error in the aperture can become quite large. In this case, the expressions in Eq. 14-73 and the patterns shown in Figure 14-38 are no longer applicable. Such a horn nevertheless exhibits axial symmetry and low cross-polarization in its radiation pattern, but pattern calculation becomes more complex, involving spherical rather than cylindrical hybrid modes.^{40-41,45} Note that the slots

are customarily cut perpendicularly to the horn wall, as in Figure 14-40, rather than to the horn axis, as in Figure 14-36a. Under balanced hybrid conditions, the radial dependence of Eq. 14-64 is now approximately $J_0(2.4048\theta/\theta_f)$, which can be used as a basis for initial design. The term *scalar horn* was coined³⁴ to emphasize the fact that the far-zone field can be represented by a single scalar quantity under ideal conditions when there is no cross-polarization. Although this is equally true for the small-flare horn, the term *scalar horn* is usually reserved for the wide-flare case.

A set of universal patterns for scalar horns under near-balanced hybrid conditions obtained by Thomas³⁹ is shown in Figure 14-41. Note that the abscissa is now θ/θ_f , where θ is the far-field polar angle and θ_f is the semi-flare angle of the horn. Since θ/θ_f is independent of frequency, it is clear that the scalar-horn patterns are almost frequency-insensitive. This is in direct contrast to the small-flare-horn case, in which beamwidth is proportional to aperture size ka . The transition is determined by the value of $\Delta = a \tan(\theta_f/2)$. For $\Delta < 0.4\lambda$ the beamwidth is controlled mainly by ka and is, therefore, frequency-dependent. As Δ increases beyond 0.5λ , this dependence is reduced, and when Δ exceeds about 0.75λ , the pattern is almost independent of Δ and hence aperture size, as shown in Figure 14-41; this effect is termed *gain saturation*. These patterns apply for values of θ_f up to about 70° . As Δ increases beyond 0.75λ , the top of the beam flattens and the skirts become steeper. At $\Delta = 1.25\lambda$, there is actually a slight dip in the axial direction. Figure 14-42 shows the effect of Δ on the beamwidth of scalar horns at various power levels relative to the peak. A good rule of thumb is that the beam semi-angle at the -12 -dB level is very close to $0.8\theta_f$ while at the -15 -dB level, it is very close to $0.9\theta_f$ for any Δ between 0.75λ and 1.5λ .

The phase factor Δ has a marked effect on the location of the phase center of the horn, as can be seen in Figure 14-43. When Δ is very small, the phase center is on the horn axis very close to the plane of the aperture. As Δ increases, the phase center moves along the axis toward the throat and eventually becomes fixed at the horn apex for scalar horns with $\Delta \approx 0.7\lambda$. The bars show the effect of a frequency change of $\pm 10\%$. The beam efficiency of scalar horns is high. When $\Delta > 0.75\lambda$, the beam efficiency at the -10 -dB level is greater than 87%, and at the -20 -dB level, it is greater than 98%. This is a valuable asset when such a horn is used as a feed in a reflector antenna since spillover loss will be small.

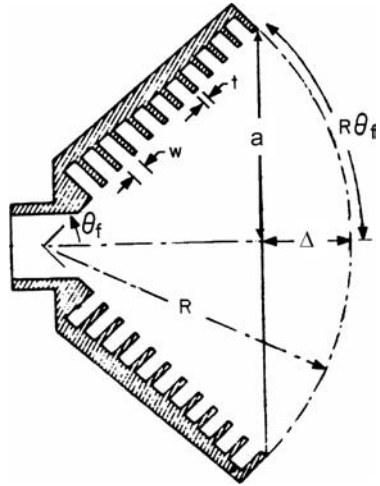


FIGURE 14-40 Wide-flare scalar horn (after Thomas³⁹ © IEEE 1978)

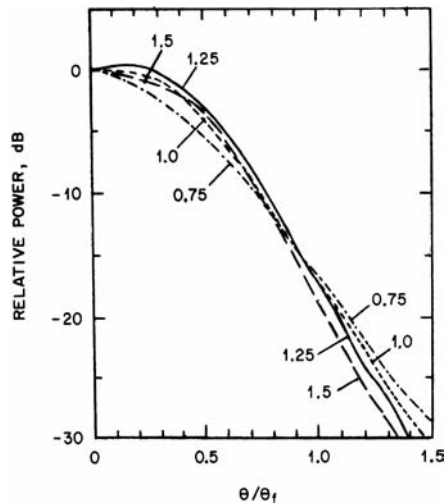


FIGURE 4.41 Universal patterns for wide-flare scalar horns ($\theta_f < 70^\circ$) with the normalized phase error Δ/λ as a parameter (after Thomas³⁹ © IEEE 1978)

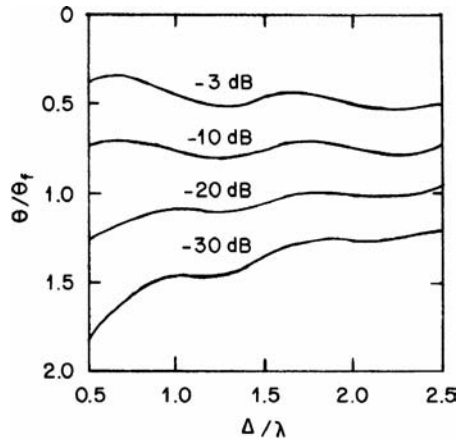


FIGURE 14-42 Effect of normalized phase factor Δ/λ on beamwidth of scalar horns at various power levels relative to beam peak (after Thomas³⁹ © IEEE 1978)

To design a scalar horn with a given beamwidth at some relative power level, say, -6 dB, one determines θ/θ_0 from Figure 14-41. It is apparent that $\theta/\theta_0 \approx 0.6$ in this case and that the parameter Δ/λ may have any value between 1.0 and 1.5. If the desired beamwidth is $2\theta = 60^\circ$, one finds that $\theta_0 = 50^\circ$, and R may lie between 2.8λ and 4.2λ .

Input Match and Slot Design

The input reflection coefficient of a corrugated horn is dominated by the reflection coefficient at the junction of the input waveguide (usually a smooth-wall circular waveguide operating in the dominant mode) and the commencement of the flare.

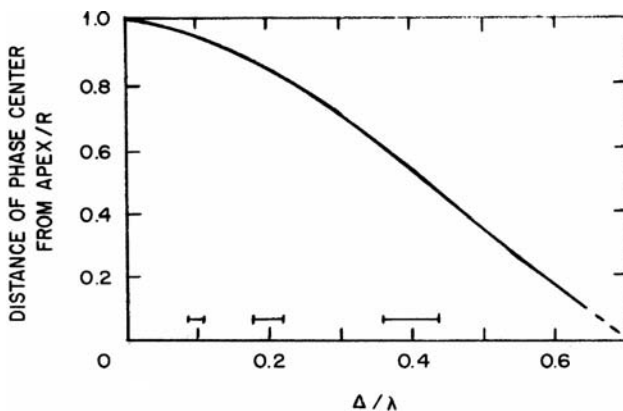


FIGURE 14-43 Distance of phase center, normalized relative to slant length R from apex (after Thomas³⁹ © IEEE 1978)

The aperture mismatch Γ_{ap} is typically < -30 dB for $a > 1.1\lambda$, and if required, can be approximated by Eq. 14-30, where the self-admittance of the HE_{11} mode under balanced hybrid conditions is

$$y_{11} = \frac{5.783}{(ka)^2} \frac{k}{\beta} \int_0^{\infty} \frac{w(2-w^2)}{\sqrt{1-w^2}} \frac{J_0^2(kaw)}{(w^2 - (2.4048/ka)^2)^2} dw \quad (14-81)$$

where $\beta = \sqrt{k^2 - (2.4048/a)^2}$. The standard design of a corrugated horn starts from the smooth wall-input circular waveguide with a section of about a wavelength long where the depth of the corrugations are gradually decreased from a half-wavelength to a quarter-wavelength deep as the horn is widened. By making the first few slots close to one-half wavelength deep, the reactance X_z will be close to zero, thus minimizing the discontinuity between the smooth conducting boundary and the corrugated wall. At the input, a linear taper can be used for the horn profile, although some other profiles, such as “hour-glass” (sine-squared), have been found more effective. To achieve the widest bandwidth operation, the first four to five slots of the input should commence by making the first slot one-half wavelength deep, then gradually reducing the slot depth until the resonant depth determined from Figure 14-36 is reached at the band-center frequency. If the depth is tapered over a distance of a wavelength or more, the match will remain good over a 1.5:1 frequency range.

The structure of a high-performance wide-band horn can be complex, as described by Thomas et al.⁴² They achieved essentially constant beamwidth, low cross-polarization, and low input reflection over a bandwidth ratio 2.2:1 with a four-section horn. The first section is a smooth-walled, profiled taper from a standard round waveguide to a large-diameter second section that acts as a converter for the HE_{11} mode. This is the most critical part of the horn because it determines return loss and strongly affects generation of unwanted EH-type modes. Ring-loaded slots are employed (Figure 14-37b) for greater bandwidth, and dimensions are optimized independently of the conical output that constitutes the fourth section. This requires the use of an intermediate, or third, section acting as a transition device. In this section, the slot depth and pitch change gradually until resonant depth is reached. There is also a gradual change in flare angle to match that of the output section. It is recommended to use five to ten slots per wavelength and the slot width-to-pitch ratio $w/(w+r) \geq 0.75$ (see Figure 14-37a). In a simpler two stage design consisting of the transition and output sections, the slot depths may be chosen empirically according to⁴⁴

$$d_n \approx \lambda \left[0.4 - \frac{(n-1)}{N_T} \left(0.4 - \frac{1}{4} \exp(2/(5ka_n)) \right) \right]; \quad n = 1, \dots, N_T \quad (14-82a)$$

$$\approx \frac{\lambda}{4} \exp(2/(5ka_n)); \quad n = N_T + 1, N \quad (14-82b)$$

where N_T is the number of slots in the transition from the input waveguide, N is the total number of corrugations, a_n is the radius of the corrugated surface at the n -th slot from the input, and λ is the wavelength at the band-center frequency. To prevent generation of the unwanted EH_{11} mode at the junction of the smooth and slotted guides, the reactance X_z should not be positive. Hence, it is advisable to ensure that the first slot is one-half wavelength deep at the highest frequency in the band. The methodology just described was used to achieve a wideband match for the horn pictured in Figure 14-1d with patterns given in Figure 14-39.¹³⁶ Wider frequency bands can be attained with even greater attention paid to this transition section. One approach that has proven particularly effective is to use ring-loaded slots, which are short-circuited coaxial waveguide slots (see Figure 14-37b).

Constant-width slots present a surface reactance that varies rapidly with frequency and is capacitive over a narrow frequency range. Ring-loaded slots present a capacitive reactance over a considerably wider bandwidth and, in addition, the frequency where $X_Z = \infty$ is lower than conventional slots. The frequency where $X_Z = 0$ depends on the slot ratio h_n/d_n (Figure 14-37*b*). Close to maximum bandwidth of 2.2:1 is possible when⁴² $h_n \approx 2d_n/3$. An empirical formula for the input width of slot n ($n = 1, \dots, N_T$) is⁴² $b'_n = p[0.1 + (n+1)(w'/p - 0.1/N_T)]$ where p is the pitch and w' is the outer width of the ring slot. The depth is now given by Eq. 14.82*b* over the full length of the horn (i.e. $n = 1, \dots, N$). At least five ring-loaded slots are needed for conversion from the input waveguide to conventional slots to realize best performance. James and Thomas⁴² provide further details of the design of ring-loaded slots.

By following the practices just described, it should be possible to obtain an input reflection coefficient $\Gamma_{in} < -18$ dB over a 1.5:1 frequency range in a scalar horn with constant width slots and to do considerably better in a small-flare horn over a narrower frequency range.

Other Corrugated Horns

The conventional transversely corrugated horn with a linear taper is just one type of such horns. Several other varieties in use today are described in the following sections and in Section 14.5, under the headings “Elliptical Horns” and “Horns with Hard Surfaces.”

Compact and Profiled Corrugated Horns The small-flare-angle horn, described earlier, tends to be quite long because Δ , the spherical wave parameter at the aperture, normally should not exceed about $\lambda/4$. Furthermore, while it retains good pattern symmetry over a relatively wide frequency range, the beamwidth varies nearly inversely with frequency. Therefore, it yields high efficiency only over a narrow frequency band. The wide-flare, or scalar, horn achieves nearly constant beamwidth and good pattern symmetry over a wide frequency range (1.5:1 or more) as long as the parameter Δ exceeds about $\lambda/2$. However, it has an aperture diameter considerably larger (by 20 to 25 percent) than the small-flare horn, and this can cause excessive blocking (or the small-flare horn may be too lengthy) when it is used as a feed for a reflector.

These shortcomings are mitigated in the compact, profiled corrugated horn in which the flare angle is not constant. Thus, the horn has an axial profile that is curved rather than following a simple linear taper. Although not essential except in wide-band operation, an input section with ring-loaded slots to effect conversion from the TE_{11} to the HE_{11} mode, followed by a specially chosen profiled section, can achieve a satisfactory performance over a frequency range of 2.4:1 from a horn significantly smaller than the conventional scalar horn. Since the cross-polarization level is not high and beamwidth and phase-center movement are not severe, this compact horn can be recommended as a feed whenever space or weight is at a premium.

As an illustration of a simple compact horn for an offset-fed reflector (half-cone angle 45°), a 12-slot corrugated horn that was designed by the methods of Thomas et al⁴² for the frequency band 10.5 to 14.5 GHz is shown in Figure 14-44. Although the number of corrugations is less than optimum, performance is satisfactory for the application. The input waveguide radius is 9.525 mm, and the aperture radius is $a = 28.92$ mm. Details of the slots are given in Table 14-6. Note that two 1-mm length uniform steps are used to represent the small taper between the slots. The experimental radiation patterns and the ones obtained with a mode-matching method are in excellent agreement and a typical result is shown in Figure 14-45 in the far-field region of the horn. At 12.5 GHz, the peak cross-polarization is about -30 dB and the gain is 15.86 dBi while the input reflection coefficient $\Gamma_{in} < -22$ dB.

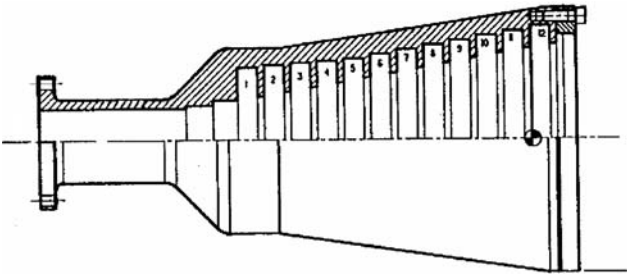


FIGURE 14-44 12-slot compact corrugated horn (input waveguide radius 9.525 mm, aperture radius $a = 28.92$ mm)

TABLE 14-6 Transverse Sections Used to Represent the 12-slot Compact Horn Profile in Figure 14-44 from Input (Section 1) to Aperture (Section 39)

Section No.	Inner Radius (mm)	Length (mm)	Section No.	Inner Radius (mm)	Length (mm)
1 (input)	9.525	45.0	21	20.46	1.0
2	10.445	8.0	22 (slot 7)	27.25	6.0
3	12	7.5	23	21.5175	1.0
4 (slot 1)	22.05	6.0	24	21.87	1.0
5	13.0575	1.0	25 (slot 8)	28.63	6.0
6	13.41	1.0	26	22.9275	1.0
7 (slot 2)	22.67	6.0	27	23.28	1.0
8	14.4675	1.0	28 (slot 9)	30.01	6.0
9	14.82	1.0	29	24.3375	1.0
10 (slot 3)	23.29	6.0	30	24.69	1.0
11	15.8775	1.0	31 (slot 10)	31.39	6.0
12	16.23	1.0	32	25.7475	1.0
13 (slot 4)	23.9	6.0	33	26.1	1.0
14	17.2875	1.0	34 (slot 11)	32.78	6.0
15	17.64	1.0	35	27.1575	1.0
16 (slot 5)	24.52	6.0	36	27.51	1.0
17	18.6975	1.0	37 (slot 12)	34.17	6.0
18	19.05	1.0	38	28.5675	1.0
19 (slot 6)	25.88	6.0	39 (aperture)	28.92	1.0
20	20.1075	1.0			

The Gaussian profile described by Eq. 14-60 is also useful for the output taper of a compact corrugated horn. It is assumed there are five to ten corrugations per wavelength of depth equal to a quarter-wavelength at the band-center frequency. The middle profile between the input section and the output profile can take on a variety of shapes.

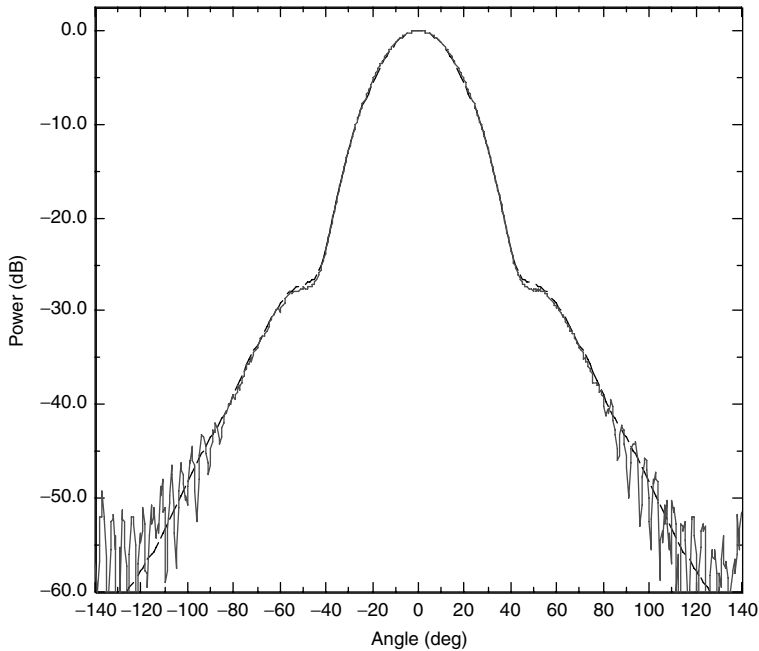


FIGURE 14-45 *H*-plane radiation pattern of the 12-slot conical corrugated horn at 12.5 GHz and a range distance of 710 mm. Solid line is measurement and dashed line is calculated result.

It can consist of several short linear sections,^{45,65} a \sin^2 profile,^{44,67} or Gaussian.^{66,139} However, the most compact horns are obtained by designing the complete profile from the input waveguide to the aperture by using optimization methods such as described in Section 14.3 under the subheading “Multimode, Stepped, and Profiled Circular Horns.” The result of one such design for a feed application is shown in Figure 14-46.⁶⁸

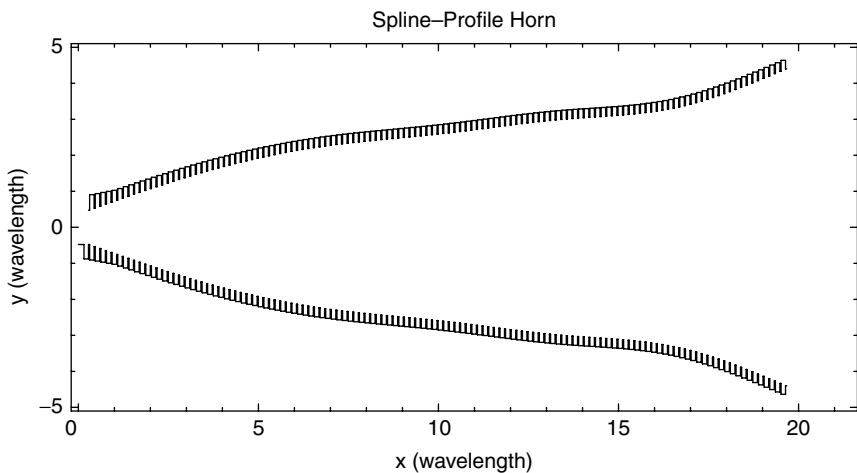


FIGURE 14-46 Spline-profile compact corrugated horn (after Granet and Bird⁶⁸ © ANTEM 2002)

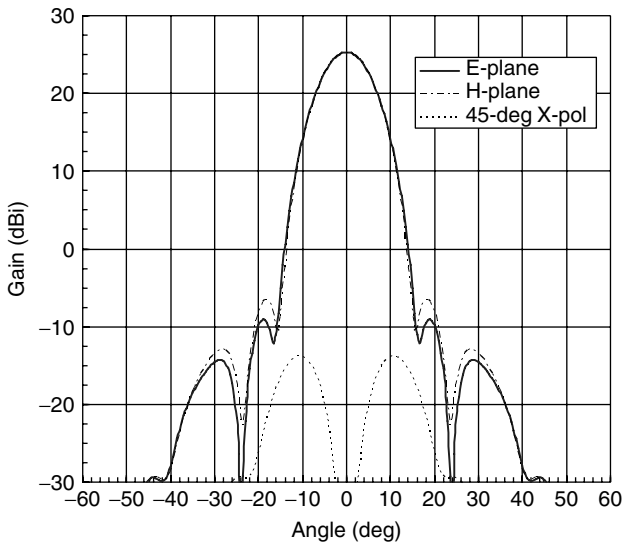


FIGURE 14-47 Radiation patterns of a spline-profile corrugated horn at 11.7 GHz

The aperture radius is $a = 85.8$ mm (4.4λ), and it is 500-mm long (19.5λ), which is a greater than 25% reduction in length from a design with a \sin^2 profile, while achieving a similar performance. The radiation pattern of this horn is given at 11.70 GHz in Figure 14-47 where the gain is 25.3 dBi. This gain is equivalent to an ideal optimum-gain corrugated horn with $a = 4.61\lambda$ and $L = 22\lambda$. However, it is emphasized that the profiled horn has significantly lower first sidelobes, peak cross-polarization level, and input reflection coefficient. The input reflection coefficient is < -25 dB over the 11.7- to 12.25-GHz frequency range. Typical for this type of design, the phase center is reasonably constant (this attribute can be included in the optimization if required) and, in this case, is situated 100 mm from the aperture inside the horn. All computer simulations were performed with a mode-matching method.

Multi-hybrid Mode Horns Additional beam shaping is possible in a corrugated horn by generating higher-order hybrid modes along with the HE_{11} . The desired higher-order modes are the HE_{1n} type ($n = 2, \dots$), all of which have axially symmetric patterns that are free of cross-polarization under near-balanced conditions, i.e., large aperture and $\gamma \rightarrow 1$. However, the higher-order modes have different dispersion characteristics, and this causes such horns to be quite frequency-sensitive.

The two-mode horn utilizes only the HE_{11} and the HE_{12} modes, both of which can be generated by a step discontinuity in the throat of the horn. The smooth-wall input guide supports only the TE_{11} mode, so that $1.84 < ka_w < 3.83$. In the corrugated region beyond the step, the radius is a_0 , and it must be large enough to support the HE_{12} as well as the HE_{11} mode but not so large as to support the HE_{13} mode. This requires $5.33 < ka_0 < 8.54$. Thomas⁷⁸ states that a significant improvement in the match near cutoff can be obtained by a gradual tapered increase in the diameter of the input guide near the step. This does result, however, in a slight reduction in the relative power of the HE_{12} mode for a given diameter $2a_0$ in the corrugated region. The HE_{12} -mode power relative to that of the HE_{11}

mode increases as the ratio a_0/a_w increases; unfortunately, a small amount of the EH_{12} mode is also produced. The presence of this mode is undesirable since it causes a significant increase in cross-polar radiation.

The radius a_0 in the corrugated section should be chosen so that ka_0 is close to 8.54, the cutoff for the HE_{13} mode at the uppermost frequency in the band of interest. If the mismatch at the upper frequency is then too large, the radius a_0 must be decreased somewhat. A compromise also must be made in regard to the TE_{11} input-guide diameter. Large values of input waveguide (ka_w near 3.83) lead to low-input reflection and a low level of the unwanted EH_{12} mode but may not result in sufficient power in the HE_{12} mode. Values between 2.2 and 3.0 for the ratio a_0/a_w have provided satisfactory results.

Because the two modes have different propagation constants, the two-hybrid-mode horn must usually incorporate a constant-diameter ($2a_0$) phasing section before it flares to its final aperture diameter. This section ensures that the modes are in the correct phase at the aperture and is exactly analogous to the use of a phasing section in the dual-mode conical horn of Figure 14-24.

The length of the corrugated flare section to the aperture is such as to antiphase the main lobes of the HE_{11} and HE_{12} radiation patterns; this creates a dip in the resultant pattern in the direction $\theta = 0$ (see Figure 14-48). The phase relationship will change with frequency and so, of course, will the pattern shape. This horn⁷⁹ was designed as a highly efficient feed for a prime-focus reflector antenna having a semi-cone angle of 63° . Similar feeds described by Thomas⁷⁸ have bandwidths of a few percent. The relative power levels in the two modes are about equal.

The pattern in Figure 14-48⁷⁹ has desirable characteristics in earth-coverage use from a satellite in a high-altitude circular orbit because of its rapid decay on the edge-of-coverage.

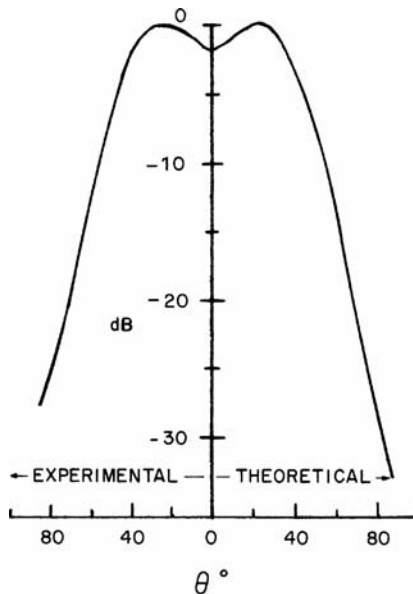


FIGURE 14-48 Experimental and theoretical patterns of a two-hybrid-mode horn with $ka_0 = 8.17$ (After Vu and Vu⁷⁹ © IEEE 1970)

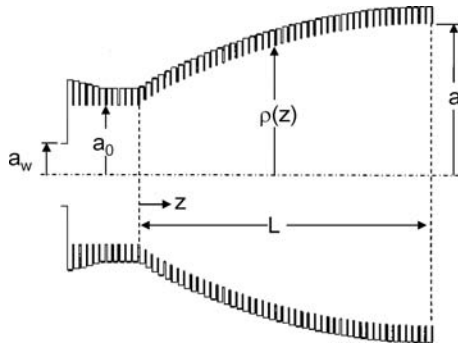


FIGURE 14-49 The bowl-shaped multimode corrugated horn

A two-mode horn of this kind was designed and tested⁸² for use on Navstar satellites of the Global Positioning System, orbiting at an altitude of 20,200 km. It has a length of 14.7 cm and an aperture diameter of 4.29 cm, which is almost exactly twice that of an optimized single-mode horn. The gain at edge of coverage of 14.3° is a minimum of 15.7 dB over a 2-GHz range in the 44-GHz band. This is 1.7 dB higher than the gain of the optimized single-mode horn at edge-of-coverage, a considerable improvement.

While the previous description has been in terms of two modes only, in reality the horns are capable of supporting more higher-order modes, and, therefore, such horns are more accurately described as *multi-hybrid-mode corrugated horns*. R.F. Thomas and coworkers⁸⁰ developed a dual-hybrid mode horn with very low sidelobes that is ideal as a feed for low-noise applications or on a satellite to provide global coverage from a geostationary satellite (half-power beamwidth of 17.4°). While the electrical performance is satisfactory (e.g., low sidelobes), this horn is very long ($\sim 30\lambda$). An alternative horn that achieves similar performance but is significantly shorter is the “bowl-shaped” horn shown in Figure 14-49.⁸³ This consists of a uniform circular waveguide input, which supports only the TE_{11} mode, a step from smooth-wall to corrugated waveguide, which creates the HE_{1n} modes, and a smooth flare transition, which serves to generate more hybrid modes and to take them from the mode generator to an aperture of the right size to give the required beamwidth. The radius defining the profile of the flare section is⁸³

$$\rho(z) = a_0 + (a - a_0) \sin^p \left(\frac{\pi z}{2L} \right) \quad (14-83)$$

where a_0 is the radius of the mode generator, a is the aperture radius, L is the length of the flare section, and z is the distance along the flare ($0 \leq z \leq L$). In its application to a global-coverage horn from geostationary orbit, an optimum index p of the profile was found to be ~ 0.8 . The mode generator section is designed with a radius of $a_0 \approx 2.2a_w$, where a_w is the radius of the waveguide exciter. When choosing a suitable multi-hybrid-mode design, the best results achieved for the application (first sidelobes < -35 dB and cross-polarization < -40 dB at the center frequency) were for a horn with an aperture diameter of 4.6λ and a total length (including mode generator) of $\sim 5.8\lambda$. The radiation pattern at the center frequency is shown in Figure 14-50. At this frequency, the measured far-field gain of this horn was identical to the calculated value of 20.85 dBi. If still shorter horns are demanded, then there will inevitably be some trade-off in low sidelobe-level performance, although this may not be too severe. Sidelobe levels < -32 dB could be achieved in the far-field of a horn about $\sim 4.5\lambda$ long while ~ -40 dB sidelobes are possible with a $\sim 6\lambda$ -long horn.

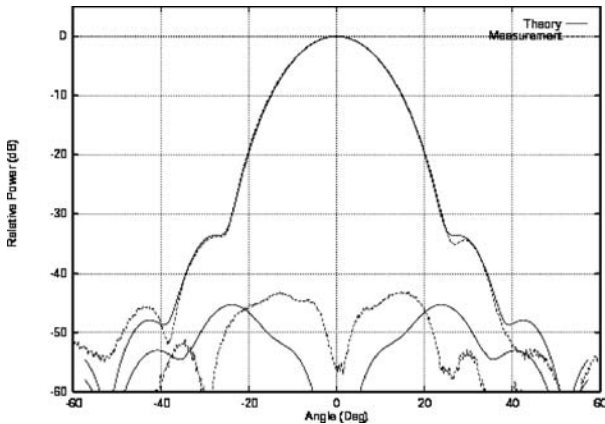


FIGURE 14-50 The 45°-plane measured and theoretical Fresnel zone radiation patterns of a bowl-shaped horn with aperture radius $a = 4.6\lambda$ at a distance of 51λ from the aperture

The return losses of the multi-hybrid-mode horn designs are typically > 25 dB and gain is > 20 dBi across a frequency band of $\sim 2.5\%$.

This horn is described as multimode because of the significant number of waveguide modes used to represent the aperture field. It was analyzed with many cylindrical waveguide sections using a mode-matching approach. The power (%) in the highest level modes making up the aperture field involves principally components TE_{11} (76.47%), TM_{11} (21.38%), TE_{12} (1.95%), TM_{12} (0.03%), and TE_{13} (0.05%).⁸³

Rectangular Corrugated Horn These horns have the advantage that patterns can be designed for applications requiring elliptical beam shapes with low cross-polarization. Sectoral and pyramidal horns excited by the TE_{10} mode in a rectangular waveguide may also benefit from the use of corrugations in the broad walls of the horn. The effect of the corrugations is to make the magnetic field small or zero at the electric walls, analogously to the vanishing electric field at the magnetic walls. As a result, the E -plane beamwidth is broadened, and the sidelobes are greatly suppressed.

In a square-aperture horn with corrugations in the electric walls, the E - and H -plane patterns will be very nearly identical, at least to the -20 -dB level, and this can be achieved over nearly a 2:1 frequency range. However, it is restricted to a single polarization, namely, when the E -plane is orthogonal to the corrugated walls.

Rectangular horns with corrugations on four faces^{7-10,41} have been designed with nearly identical patterns in the principal planes for circular polarized radiation with low sidelobes. As with all corrugated horns, the transition from smooth-wall input waveguide to the corrugations is made with a slot depth of a half-wavelength to a quarter-wavelength for a good reflection match. Improved pattern control can be obtained by filling the corners with metal where the corrugations meet.^{8,10} The bandwidth and poor cross-polarization of the rectangular corrugated horn is due to the generation of a higher-mode slow wave, which is predicted in the rigorous analysis of the horn.¹⁰ The bandwidth of a rectangular corrugated horn is typically less than 40%, which is somewhat less than the bandwidth of greater than 70% that can be achieved with a circular corrugated horn. Detailed simulation is necessary to obtain the appropriate mode combination in the aperture for the general case, and the use of accurate simulation software is recommended for any design.

14.5 MISCELLANEOUS HORN STRUCTURES

The horns described in the previous sections are the types most frequently encountered. However, there are many other types of horns, and a sample is described in the following section along with some accessories, such as lenses, that can help improve performance in certain situations.

Elliptical Horns

The beams radiated by an elliptical horn have much in common with smooth-wall rectangular horns, except that now a greater variety of tapers are possible in the E -plane.⁹¹ The dominant mode of the air-filled elliptical horn in which the transverse electric field is parallel to the minor axis (see Figure 14-51) of the ellipse is the TE_{c11} even-order mode. Its orthogonal counterpart, having a transverse electric field parallel to the major diameter, is called an *odd-order mode*. It has been shown that for each even mode there exists an odd mode having the same propagation constant. Furthermore, the two modes will have identical, but non-axially symmetric radiation patterns with orthogonal polarizations. It then follows that if the two orthogonal modes are excited in such a way as to have equal amplitudes and a 90° phase difference, the radiated field will be everywhere circularly polarized.

The description of the radiated field is similar to that for a rectangular horn (see Section 14.2), and some of the results can be carried directly over to elliptical horns. The complication is that the radiated fields for the TE_{c11} even-order mode are given in terms of Mathieu functions.¹⁴⁰ Although more accurate simulation by numerical software bypasses the need of these functions. There are computer subroutines for Mathieu functions¹⁴¹ and some symbolic mathematical software (e.g., Mathematica¹⁴²) include them as embedded functions. When a horn of axial length h and aperture with semi-major and minor dimensions a and b is fed from an elliptical waveguide with semi-major and minor dimensions a_w and b_w , the fields of the dominant mode TE_{c11} mode are

$$E_\theta(r, \theta, \phi) = F_\theta (N_x \sin \phi - N_y \cos \phi) \quad (14-84a)$$

$$E_\phi(r, \theta, \phi) = F_\phi (N_x \cos \phi + N_y \sin \phi) \quad (14-84b)$$

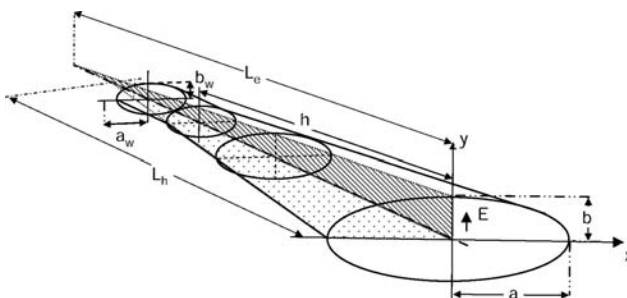


FIGURE 14-51 Geometry of an elliptical horn

where

$$N_x = c_0 \int_0^{\xi_0} \int_0^{2\pi} \left[Ce_1(\xi, q) ce_1'(\eta, q) \sinh \xi \cos \eta + Ce_1'(\xi, q) ce_1(\eta, q) \cosh \xi \sin \eta \right] \\ \times \exp \left[jk \left(\sin \theta (x' \cos \phi + y' \sin \phi) - \frac{1}{2} \left(\frac{x'^2}{L_h} + \frac{y'^2}{L_e} \right) \right) \right] d\eta d\xi \quad (14-85a)$$

$$N_y = c_0 \int_0^{\xi_0} \int_0^{2\pi} \left[Ce_1(\xi, q) ce_1'(\eta, q) \cosh \xi \sin \eta - Ce_1'(\xi, q) ce_1(\eta, q) \sinh \xi \cos \eta \right] \\ \times \exp \left[jk \left(\sin \theta (x' \cos \phi + y' \sin \phi) - \frac{1}{2} \left(\frac{x'^2}{L_h} + \frac{y'^2}{L_e} \right) \right) \right] d\eta d\xi \quad (14-85b)$$

where $c_0 = jk\ell E_0 e^{-jk\ell/2\pi}$; $x' = \ell \cosh \xi \cos \eta$; $y' = \ell \sinh \xi \sin \eta$; $\ell = ae = \sqrt{a^2 - b^2}$; a and b are the aperture semi-major and minor axes lengths; e is the eccentricity; $ce_1(\eta, q)$ and $Ce_1(\xi, q)$ are the normal and modified even Mathieu functions of order one, respectively, on which a prime denotes the first derivative with respect to the first argument; and q is the first zero of $Ce_1'(\xi_0, q) = 0$, where $\cosh \xi_0 = 1/e$, and the cutoff frequency of the dominant mode is $k_{c11} = 2\sqrt{q}/\ell$. The parameters L_e and L_h are the axial lengths in the E - and H -planes (see Figure 14-51), in common with rectangular horns, and are given by equivalent equations to Eq. 14-13. Equation 14-85 must be integrated numerically to obtain the far-fields from Eq. 14-84, which are found to furnish useful design information for directive horns in the vicinity of the main beam with $F_\theta = 1$ and $F_\phi = \cos \theta$, particularly for the H -plane radiation pattern within the limitations of the quadratic phase approximation (e.g., dominant mode, flare angles $< 35^\circ$).

Normally, the patterns in the two principal planes are different for an elliptical horn, usually with a shoulder in the E -plane at about ~ -10 dB and a monotonically decreasing H -plane pattern. Lower sidelobes can be obtained in the E -plane by varying the eccentricity along the horn's length (there is a different quadratic phase factor than in Eq. 14-85). With some geometries, the beam can be made almost symmetrical.⁹⁹

The design procedure for elliptical horns is identical to the methodology described for rectangular horns in the section, "Pyramidal Horns." Optimum-gain elliptical horns are also possible. To a good approximation $a \approx \sqrt{0.94 L_h \lambda}$ and $b \approx \sqrt{0.65 L_e \lambda}$, and the aperture efficiency is $\sim 50\%$ for this family of horns. This means the solution of Eq. 14-25 for an optimum-gain rectangular horn given by Eq. 14-26 is applicable also to an optimum-gain elliptical horn, wherein $\alpha_1 = 0.94\lambda$, $\beta_1 = 0.65\lambda$, $g_1 = 2\pi^2$, and it is understood that a , a_w , and so on, now refer to the corresponding ellipse semi-axes (see Figure 14-51). The H -plane pattern of the optimum horn family has low sidelobes while in the E -plane, the first sidelobe is about -10 dB (compared with -9 dB for a pyramidal horn and ~ -8.5 dB for a conical one).

Elliptical corrugated horns have also proven useful,⁹³⁻⁹⁵ and moderately wideband performance (up to 40%) has been reported through careful design of the transition from the feeder waveguide to the aperture.⁹⁵ The copolar radiation pattern at the balanced hybrid condition can be calculated approximately for initial design purposes from an aperture field for a round corrugated horn given by Eq. 14-64 with a quadratic phase factor¹⁴³ plus information given in the "Input Match and Slot Design" section in Section 14.4 to determine the slot depth and pitch. It is assumed that in the principal planes of the elliptical horn (i.e., the planes containing the major and minor axes) the radiation patterns will be approximately the same as for those of a round horn of corresponding aperture diameters. If ka and kb are moderately large, it is permissible to use constant depth in any one slot, rather than one that varies with the angle ϕ . This greatly simplifies construction of the horn.

Dielectric loading may be used in lieu of corrugations to generate elliptically shaped beams with good circular polarization over a broadband.^{96–98} Both elliptical- and rectangular-aperture horns have been used successfully for this purpose.

Ridged Horns

As we have seen, up to about 2.4:1, bandwidth can be obtained with a corrugated horn by special attention to the input match. When wider bandwidths than this are required, possibly extending over several octaves or even frequency decades, other approaches are needed that ultimately depend on attaining wideband transition from other components such as the feeder. The input match is the usual parameter used to define the operation of ultra wideband, although constant gain performance is sometimes used. The horn patterns vary significantly over wide frequency ranges; nevertheless, a return loss of greater than 10 dB can be achieved over several octaves.

The most common wideband transition is related to the ridged waveguide—double-ridged for single polarization and quad-ridged for dual polarization.²⁴ Waveguides, such as ridged waveguide, with re-entrant cross-sections, have wide frequency separation between the cutoff of the fundamental TE mode and the next higher-order mode. This is due to the constriction between the ridge and the two waveguides on either side of it. Any of the horn types described previously can be used as the radiating element. Rectangular,^{1,24–27} conical,^{33,85–88} and dielectric-loaded horns are the most usual (for example, a quad-ridged conical horn is illustrated in Figure 14-52). Historically, the work of Walton and Sundberg²⁴ on the double-ridge rectangular horn was important and is still in use today for the design of such horns with over a decade frequency range. The aim is to taper the ridges to make a smooth transition from the input (usually a coaxial connection) to the aperture to achieve a good match to the aperture. This is done by ensuring the ridges taper as the horn flares outward. For dominant mode horns, it is common practice to keep the cutoff frequency of the dominant mode approximately constant as the ridged horn flares. It is important that everywhere along the horn the geometry can support the dominant mode at the lowest frequency. Otherwise the mode can become trapped, and although tunneling can occur, there will be significant input mismatch and gain will suffer. The ridge is rolled off from about a quarter of the way down the flare using a common profile such as a \sin^2 or exponential taper (see Figure 14-53). If quad-ridges are used, it is vital to maintain symmetry between the ridges, otherwise cross-polarization will suffer. Mechanical accuracy and good contact of the ridges to the walls is necessary to achieve the widest range of frequencies and best performance. Double-ridge horns can be designed approximately by the methods outlined in Bruns et al.²⁶ and the performance tuned by electromagnetic software. A double-ridge horn simulated in this

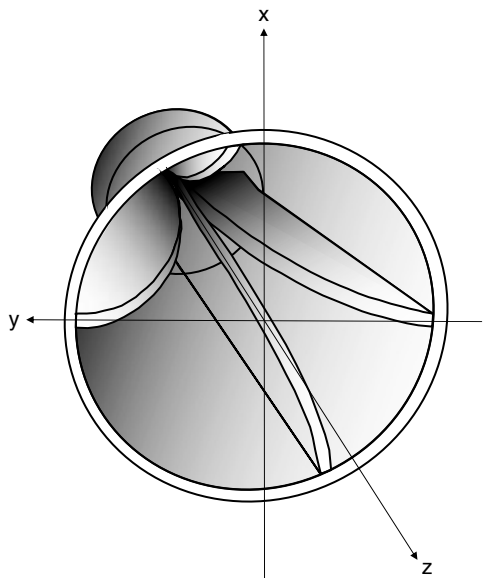


FIGURE 14-52 Quad-ridged conical horn

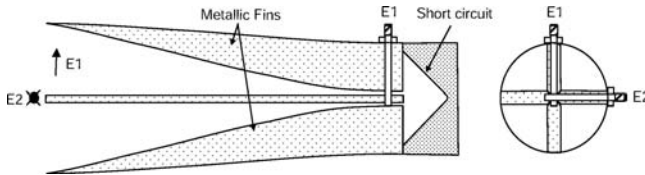


FIGURE 14-53 Sectional perspectives of the input to a quad-ridged horn. E1 and E2 refer to the two polarizations.

way gave good agreement with measurement over part of the 1–18-GHz operational range.²⁶ This work showed that at high frequencies, the main beam splits into multiple lobes. A quad-ridged horn with no sidewalls overcomes this problem and operates successfully over the full frequency range.²⁷

The widest bandwidth with a ridge waveguide is made possible when the input ridges are very close together. This narrow spacing of the ridges ultimately determines the lowest frequency for a given geometry. It also means that wideband horns are difficult to construct for the millimeter-wave band. In addition, a reasonable length of taper is required to give wideband performance. If the ridge is too short, mismatch and frequency performance will be poor. It is recommended that the length of the taper should be at least four times the wavelength at the lowest frequency. As a result, wideband feed horns are quite long. If higher than about a –10-dB mismatch loss can be tolerated, the horn length can be shortened significantly for a reduced frequency range. Typically a 1λ length or less horn can give a 3:1 operational bandwidth with a maximum return loss of ~6 dB.²⁵ An example of the input match of a quad-ridged conical horn⁸⁶ is shown in Figure 14-54. The bandwidth of this horn is 10:1, and the return loss across the band is greater than 10 dB.

The radiation pattern can be predicted approximately by neglecting the ridges in the aperture, although in practice the ridges have a second-order effect. The gain of a ridged-waveguide-fed horn can be estimated also from the gain of the corresponding flared geometry. For example, the measured gain of a double-ridge horn is lower than predicted by Eq. 14-15 for a pyramidal horn by 1–2.25 dB.¹⁴⁴

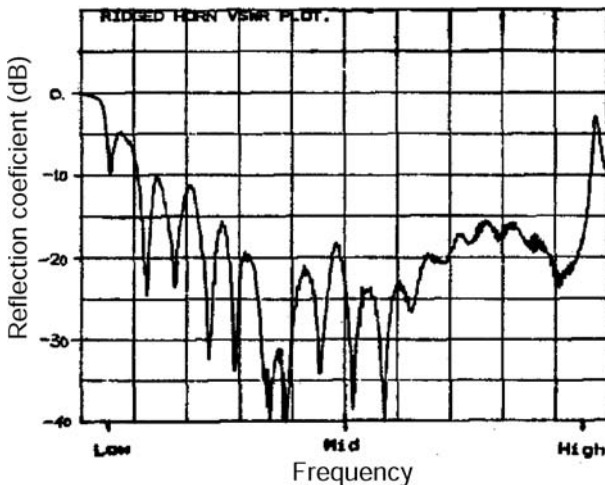


FIGURE 14-54 Measured reflection coefficient of quad-ridged conical horn (after Lai et al⁸⁶ © IEEE 1987) that covers a 10:1 bandwidth

Horns with Hard Surfaces

A horn with a hard surface condition, or hard horn, satisfies Eq. 14-72. This implies that both axial field components vanish at the horn's hard inner surface. For this to occur in a circular geometry, Eq. 14-63 requires that the radial-separation constant k_ρ must be zero and the mode must be TEM. It is then found that the transverse components become

$$e_\rho = -e_\phi = h_\rho = h_\phi = -j(1 + \gamma)/2 \quad (14-86)$$

The value of the mode-content factor γ is quite immaterial in this case compared with the soft horn. If it is zero, then the mode is a degenerate form of TM_{11} ; if it is infinite, it is a degenerate form of TE_{11} . The mode also can be regarded as degenerate HE_{11} if $\gamma = 1$; if $\gamma = -1$, however, all components vanish. The components in Eq. 14-86 show that the Poynting vector is z -directed, constant, and also independent of ρ , as befits a hard surface. Furthermore, Eq. 14-64 indicates that the aperture field is unidirectional containing only the component e_x . In the absence of spherical wave-phase error, this mode radiates with 100% aperture efficiency and -17.6 -dB first sidelobes. By comparison, the ideal soft horn has 69% efficiency, -27.5 -dB first sidelobes, and a half-power beamwidth that is 28% larger than that of the ideal hard horn of the same aperture diameter.¹³²

A horn having longitudinal corrugations approximates the hard boundary condition when the depth of corrugation is⁷⁵

$$d \approx \frac{\lambda}{4\sqrt{\epsilon_{r2} - \epsilon_{r1}}} \quad (14-87)$$

where ϵ_{r1} and ϵ_{r2} are the permittivity of the materials filling the horn and in the corrugations. Note that Eq. 14-87 is identical to Eq. 14-39 for a rectangular horn with dielectric lining on the two E -plane walls (see "Dielectric-lined Rectangular Horns," earlier in the chapter) and, therefore, this latter horn has two hard surfaces. Square horns have been realized with four longitudinally corrugated surfaces filled with dielectric.¹⁴⁵ Longitudinally aligned strips on dielectric can also meet the hard boundary condition when the dielectric thickness satisfies Eq. 14-87. Both filled corrugation and strip-loaded dielectric surfaces work satisfactorily over a relatively narrow bandwidth for any aperture size and shape. The horns tend to have higher efficiency as the aperture size increases due to the power in the wall area decreasing relative to the total power. A circular dielectric-loaded horn can also be designed to have a hard boundary surface.⁵¹ The dielectric load has an inner core of permittivity ϵ_{r1} that has a coat of higher permittivity material (ϵ_{r2}) of thickness t_2 . This coated dielectric load is placed at a uniform distance t_3 from the metallic horn wall. The spacing material has permittivity ϵ_{r3} and is usually air with symmetric dielectric supports to minimize cross-polarization. The material parameters are chosen so that $\epsilon_{r2} > \epsilon_{r1} > \epsilon_{r3}$. The coating thickness for hard surfaces is typically $t_2 < 0.15\lambda$ for $\epsilon_{r1} = 1.3$ and reduces as ϵ_{r2} is increased. The total spacing thickness $t_2 + t_3$ is also lower for high ϵ_{r2} . For example, when $\epsilon_{r1} = 1.3$ and $\epsilon_{r2} = 4$, $t_2 + t_3 \approx 0.31\lambda$. Horns designed with these surfaces have the potential for achieving pattern performance approaching a corrugated horn, although with aperture efficiencies of greater than 90%.⁵¹

Absorber-lined Horns

It has been shown that a conical horn with a lossy dielectric layer (relative permittivity $\epsilon_r = \epsilon'_r + j\epsilon''_r$) can support nearly balanced hybrid modes¹⁴⁶ when the horn diameter is large ($ka \gg 1$). The thickness of the layer and the dielectric-loss tangent should be large

enough to ensure that one-way transmission is attenuated by at least 10 to 15 dB. In this case, the eigenvalue $k_p a$ will be close to the desired value, 2.4048, and it will have a small imaginary part. As a consequence, the propagation constant also will have a small imaginary part, which means that the HE_{11} mode suffers a slight ohmic loss. There is then no upper limit to the frequency of hybrid-mode operation. The absorber thickness decreases as ϵ'_r and ϵ''_r increase, and in a horn with large ka , the field radiated from the absorber annulus will be negligible.¹⁴⁷ A nearly balanced HE_{11} mode operation was first demonstrated in a horn reflector with an aperture diameter of 217 cm by Knop et al.¹⁴⁷ An absorber dielectric with $\epsilon'_r = 1.4$ and $\epsilon''_r = 0.56$ lines the upper part of the cone and enables the horn to operate from 4 to 12 GHz, the upper limit in the frequency being set by the nature of the waveguide launcher at the throat of the horn. The ohmic loss is less than 0.5 dB over the band.

A conical horn lined with magnetically lossy material has been shown to exhibit hybrid-mode operation¹⁴⁸ but at the expense of 10 dB of ohmic loss. In a different version,¹⁴⁹ the loss was considerably reduced, being between 0.8 and 2.7 dB.

Horns with a Protruding Central Conductor

A horn with a central conductor that extends from the aperture is used to support the horn as a feed and in medical probes. Some examples of horns with an extended central conductor are the conical and rectangular horns, Cutler, hat, and cup feeds.^{21–23,59–64} The central conductor can enclose a coaxial cable or waveguide to feed the horn or contain other sensing apparatus. An example of a conical horn with a protruding conductor is shown in Figure 14-55. The diameter of the central rod influences the performance of the horn, and to retain most of the properties of the original horn, the rod radius b should satisfy $kb < \sim 2$ and $b/a < 0.25$, where a is the aperture radius (i.e., $b < 0.3\lambda$ and $a > 1.2\lambda$).⁵⁵ While this can be easily met with a coaxial cable feed, a waveguide feed may need to be dielectric-loaded to keep the rod radius as small as possible. When the conditions above are not satisfied, the H -plane pattern is narrower and the E -plane pattern wider compared with the original horn, increasing the level of cross-polarization.⁵⁵ The horn can be fed directly by the coaxial cable connected to a dipole exciter, a pair of slots, a waveguide reflector at the closed back of the horn or a folded waveguide transition. The exciter can limit the usable bandwidth to only a few percent of the operating frequency if care is not taken with the method of feeding. Wider bands are possible ($\sim 20\%$) using suitably designed excitation. Choke rings in a flange near the aperture can improve the radiation pattern symmetry of small aperture horns ($ka < 4$).⁶²

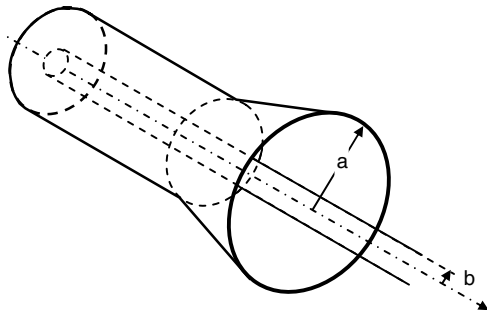


FIGURE 14-55 A conical horn with protruding central conductor

Horns with Lenses

A lens is sometimes added to a horn, usually at the aperture, to alter the radiation performance to suit the application. This is especially true for correcting phase errors and for realizing significant directivity in a confined length. A dielectric lens is usually preferred as metal plate lenses cause diffraction and depolarization that significantly change the behavior of the original horn. A lens can increase aperture efficiency, reduce sidelobes, or produce a combined phase center that is less sensitive to frequency. There are many different types of lenses for horns based on the surface profile of the inner and outer surfaces.⁴⁵ The most common type for horns is the plano-convex lens (see Figure 14-56), which has plane inner and convex outer surfaces. To simplify design, lenses are mostly circularly symmetric, although elliptical lenses have been produced for elliptical horns. The profile of a plano-convex lens is given by

$$\rho = \frac{\sin \theta (F + (\sqrt{\epsilon_r - 1}t)) + \tan \theta F (\epsilon_r - 1 - \sqrt{\epsilon_r - \sin^2 \theta})}{\epsilon_r - \sqrt{\epsilon_r - \sin^2 \theta}} \quad (14-88a)$$

$$z = \left(\frac{\rho}{\sin \theta} - \frac{F}{\cos \theta} \right) \sqrt{\epsilon_r - \sin^2 \theta} \quad (14-88b)$$

where F is the distance to the focal point, t is the lens thickness, and ϵ_r is the dielectric constant of the lens material. The wavefront from the lens is planar as long as the horn produces spherical waves, which implies that the combined phase center of the radiation is independent of frequency. The lens thickness is given by

$$t = \frac{F}{(\sqrt{\epsilon_r - 1})} \left(\sqrt{1 + \frac{1}{(2F/D)^2}} - 1 \right) \quad (14-89)$$

where D is the lens diameter. Clearly, the lens is thicker with lower dielectric constant materials, which usually means weight and cost problems in manufacturing and application. To overcome these deficiencies, the lens surface can be “zoned” so the optical path from one

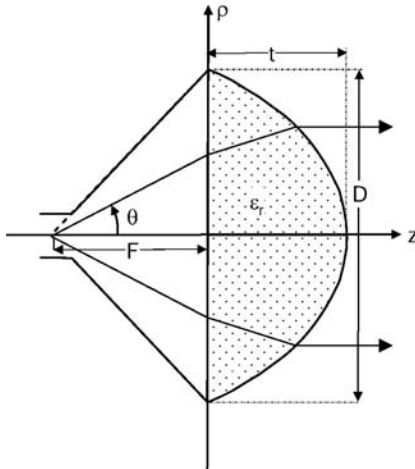


FIGURE 14-56 Geometry of horn with plano-convex lens

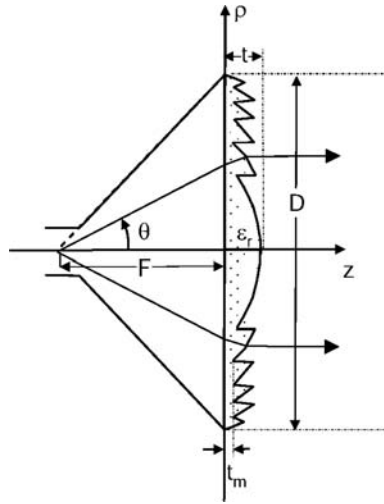


FIGURE 14-57 Geometry of horn with zoned plano-convex lens

zone to the next differs by multiples of a wavelength (Figure 14-57). The design methodology is detailed in Silver.¹⁶ Commencing at the outer edge of the lens (zone N) and moving toward the center ($n = N, N - 1, \dots, 1$), when the path length through the lens exceeds one wavelength in the material ($\lambda / \sqrt{\epsilon_r}$), the path length is reduced by one wavelength without effecting the phase of the wavefront, and this is achieved by reducing the lens profile by the corresponding amount. Zone $N - 1$ is handled the same way as are the subsequent zones until the center is reached. The maximum thickness of a zoned lens is $\sim \lambda / (\sqrt{\epsilon_r} - 1)$, although in practice a small amount of material of thickness t_m is added to t for mechanical support. A zoned plano-convex lens is shown in Figure 14-57. This time the lens properties are frequency dependent, and therefore, as well as designing for a particular horn, the frequency band should be taken into account in the lens design.

The radiation pattern of the horn and lens combination is computed approximately by means of geometric optics and aperture integration. The aperture amplitude distribution of the lens and horn is given by

$$A(\rho) = A(\theta) \sqrt{\frac{\sin \theta}{\rho} \frac{d\theta}{d\rho}} \quad (14-90)$$

where $A(\theta)$ is the amplitude distribution of the source. The radiation pattern can be obtained from Eq. 14-90 by computing the derivative of the lens profile (from, for example, Eq. 14-88) and the amplitude function of the source, such as for conical (Eq. 14-46) or corrugated horns (Eq. 14-74), and integrating over the aperture variables ρ and ϕ as described in Chapter 1.

The performance of the lens-corrected horn is limited by several factors including reflections from the lens surface and dielectric loss. Reflections for plano-convex lenses tend to be relatively small, and the dominant mismatch is due to the horn itself. Dielectric loss is strongly determined by the path through the thickest part of the lens (i.e., center). An upper estimate of lens loss is given by the plane wave attenuation of $27.3 t \sqrt{\epsilon_r} \tan \delta$ dB, where $\tan \delta$ is the loss tangent of the lens material. Although the above methodology is based on simple concepts, the agreement with measurement is usually sufficient for

most practical purposes.¹⁵⁰ For example, a hyperbolic plano-convex lens (the convex surface is facing the source compared with Figure 14-56), designed for a conical horn for a 60-GHz wireless link,¹⁵¹ gave good performance when zoned, and at the same time, the combination was highly compact. Accurate simulation software should be used to correct for second-order effects.

Dielectric Horns

Solid dielectric horns have additional contributors to the radiation than the metallic horns discussed in the previous sections. There are three basic sources of radiation that need to be included. From the input of the horn, they are the (i) transition from the source (usually a waveguide) to the dielectric, (ii) radiation from the tapered horn as the wave travels toward the aperture, and (iii) radiation from the aperture. Radiation from source (iii) normally provides the greatest contribution. The dominant mode is a hybrid HE_{11} type, and the difficulty of obtaining the field and propagation parameters of this mode, as well as the three-part radiation mechanism, make dielectric horns more complex to design than metallic horns. The directive beam is determined by the aperture and to block radiation from sources (i) and from the tapered horn (ii), the end launcher and flare can be shielded or coated with absorbing material. Solid dielectric horns with conventional geometries, such as pyramidal and conical types, have been designed with some success.¹⁰²⁻¹⁰⁸ Weight can be reduced by hollowing the dielectric as has been described in Palit¹⁰⁷ for a pyramidal horn; the hollowing also lowers the E -plane sidelobes relative to the H -plane. The computed radiation for a dominant mode approximation with the phase-shift variation from the input to the aperture included gave good agreement with experiment. Less conventional feed structures can be employed with dielectric horns, such as a two-wire line and launcher plates,¹⁰⁸ which when applied to a pyramidal dielectric horn, the FDTD method predicts constant input impedance over a 120% bandwidth.

Acknowledgment

The authors are grateful for the assistance of C. Granet for the results using CST Microwave Studio.

REFERENCES

1. A. W. Love (ed.), *Electromagnetic Horn Antennas* (New York: IEEE Press, 1976).
2. C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd Ed. (New York: John Wiley & Sons, 1996).
3. T. S. Chu and R. A. Semplak, "Gain of Electromagnetic Horns," *Bell Syst. Tech. J.*, vol. 44 (March 1965): 527-537.
4. E. V. Jull, "Errors in the Predicted Gain of Pyramidal Horns," *IEEE Trans. Antennas Propag.*, vol. AP-21, no. 1 (January 1973): 25-31.
5. W. M. Truman and C. A. Balanis, "Optimum Design of Horn Feeds for Reflector Antennas," *IEEE Trans. Antennas Propag.*, AP-22 (1974): 585-586.
6. R. E. Lawrie and L. Peters, Modifications of Horn Antennas for Low Sidelobe Levels," *IEEE Trans. Antennas Propag.*, AP-14 (1966): 605-610.
7. G. H. Bryant, "Propagation in Corrugated Waveguides," *Proc. IEE*, vol. 116 (1969): 203, 213.
8. R. Baldwin and P. A. McInnes, "A Rectangular Corrugated Feed Horn," *IEEE Trans. Antennas Propag.*, AP-23 (1975): 814-817.
9. C. Dragone, "A Rectangular Horn of Four Corrugated Plates," *IEEE Trans. Antennas Propag.*, AP-33 (1985): 160-164.

10. L. C. da Silva and S. Ghosh, "A Comparative Analysis Between Rectangular Corrugated Waveguides with Plain and Corner Filled Corrugations," *IEEE AP-S International Symposium* (1990): 976–979.
11. G. N. Tsandoulas and W. D. Fitzgerald, "Aperture Efficiency Enhancement in Dielectrically Loaded Horns," *IEEE Trans. Antennas Propagat.*, AP-20 (1972): 69–74.
12. E. Lier and R. Landes, "High Gain Linearly Polarized Dual Band Horn for Multi-feed Reflector Antenna," *Electron. Lett.*, vol. 39 (2003): 1230–1232.
13. S. B. Cohn, "Flare-angle Changes in a Horn as a Means of Pattern Control," *Microwave J.*, vol. 13 (October 1970): 41–46.
14. A. Bhattacharyya and G. Goyette, "A Novel Horn Radiator with High Aperture Efficiency and Low Cross-polarization and Applications in Arrays and Multibeam Reflector Antennas," *IEEE Trans. Antennas Propagat.*, 52 (2004): 2850–2859.
15. T. S. Bird and C. Granet, "Optimized Design of High-efficiency Rectangular Horns for Arrays," *2004 URSI Electromagnetic Theory Symposium*, Pisa, Italy (May 23–27, 2004): 963–965.
16. S. Silver, *Microwave Antenna Theory and Design* (New York: McGraw-Hill, 1949).
17. P. A. Jensen, "A Low-noise Multi-mode Cassegrain Monopulse Feed with Polarization Diversity," *Northeast Estro. Res. and Eng. Meeting* (November 1963): 95–95.
18. N. Amitay and M. J. Gans, "Design of Rectangular Horn Arrays with Oversized Aperture Elements," *IEEE Trans. Antennas Propagat.*, AP-29 (1981): 871–884.
19. K. M. Lee and R. S. Chu, "Design and Analysis of a Multimode Feed Horn for a Monopulse Feed," *IEEE Trans. Antennas Propagat.*, AP-36, (1988): 171–181.
20. T. S. Bird, "Mode Matching Analysis of Arrays of Stepped Rectangular Horns and Application to Satellite Design," *IEE International Conference on Antennas and Propagation (ICAP)* (1991): 849–852.
21. C. C. Cutler, *Directional Microwave Antenna* (1947): US Pat. 2, 422, 184.
22. A. Chlavin, "A New Antenna Feed Having Equal E- and H-Plane Patterns," *IRE Trans. Antennas Propagat.*, vol. AP-2 (July 1954): 113–119.
23. N. Bui-Hai, *Antennes Micro-ondes* (Paris: Masson, 1978).
24. K. L. Walton and V. C. Sundberg, "Broadband Ridged Horn Design," *Microwave J.* (1964) 96–101
25. S. A. Soroka, "A Physically Compact Quad Ridge Horn Design," *IEEE AP-S International Symposium Digest* (1986): 903–906.
26. C. Bruns, P. Leuchtmann, and R. Vahldieck, "Analysis and Simulation of a 1-18 GHz Broadband Double-ridge Horn Antenna," *IEEE Trans. Electromag. Compatibility*, vol. 45 (February 2003): 55–60.
27. V. Rodriguez, "An Open-boundary Quad-ridged Guide Horn Antenna for Use as a Source in Antenna Pattern Measurement Anechoic Chambers," *IEEE Antennas and Propagation Magazine*, vol. 48 (April 2006): 157–160.
28. G. C. Southworth and A. P. King, "Metal Horns as Directive Receivers of Ultra-short Waves," *IRE Proc.*, vol. 27 (February 1939): 95–102.
29. A. P. King, "The Radiation Characteristics of Conical Horn Antennas," *IRE Proc.*, vol. 38 (March 1950): 249–251.
30. M. A. K. Hamid, "Diffraction by a Conical Horn," *IEEE Trans. Antennas Propagat.*, vol. AP-16 (September 1968): 520–528.
31. M. S. Narasimhan and B. V. Rao, "Modes in a Conical Horn: New Approach," *IEE Proc.*, vol. 118 (February 1971): 287–292.
32. M. S. Narasimhan and B. V. Rao, "Radiation from Conical Horns with Large Flare Angles," *IEEE Trans. Antennas Propagat.*, vol. AP-19 (September 1971): 678–681.
33. R. Burger, R. Beyer, and F. Arndt, "Rigorous Combined Mode-matching Integral Equation Analysis of Horn Antennas with Arbitrary Cross Section," *IEEE Trans. Antennas Propagat.*, 47 (November 1999): 1641–1648.
34. A. J. Simmons and A. F. Kay, "The Scalar Feed—A High Performance Feed for Large Paraboloid Reflectors," *IEE Conf. Publ. 21* (1966): 213–217.

35. H. C. Minnett and B. M. Thomas, "A Method of Synthesizing Radiation Patterns with Axial Symmetry," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (September 1966): 654–656.
36. V. H. Rumsey, "Horn Antennas with Uniform Power Patterns Around Their Axes," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (September 1966): 656–658.
37. B. M. Thomas, "Theoretical Performance of Prime Focus Paraboloids Using Cylindrical Hybrid Modes," *IEE Proc.*, vol. 118 (November 1971): 1539–1549.
38. P. J. Clarricoats and P. K. Saha, "Propagation and Radiation Behavior of Corrugated Feeds," *IEE Proc.*, vol. 118, parts 1 and 2 (September 1971): 1167–1186.
39. B. MacA. Thomas, "Design of Corrugated Conical Horns," *IEEE Trans. Antennas Propagat.*, AP-26 (1978): 367–372.
40. J. K. M. Jansen, M. E. J. Jeuken, and C. W. Lambrechtse, "The Scalar Feed," *Arch. Elek. Übertragung*, vol. 26 (January 1972): 22–30.
41. P. J. B. Clarricoats and A. D. Olver, *Corrugated Horns and Microwave Antennas* (London: Peter Peregrinus Ltd., 1984).
42. G. L. James and B. MacA. Thomas, "TE₁₁ to HE₁₁ Cylindrical Waveguide Mode Converters Using Ring-loaded Slots," *IEEE Trans. Microw. Theory Tech.*, MTT-30 (1982): 278–285.
43. B. MacA. Thomas, G. L. James, and K. J. Greene, "Design of Wide-band Corrugated Conical Horns for Cassegrain Antennas," *IEEE Trans. Antennas Propagat.*, AP-34 (1986): 750–757.
44. G. L. James, "Design of Wide-band Compact Corrugated Horns," *IEEE Trans. Antennas Propagat.*, AP-32 (1984): 1134–1138.
45. A. D. Olver, P. J. B. Clarricoats, A. A. Kishk, and L. Shafai, *Microwave Horns and Feeds* (London: Peter Peregrinus Ltd, 1994).
46. T. Satoh, "Dielectric-loaded Horn Antenna," *IEEE Trans. Antennas Propagat.*, AP-20 (1972): 199–201.
47. E. Lier, "A Dielectric Hybrid Mode Antenna Feed: A Simple Alternative to the Corrugated Horn," *IEEE Trans. Antennas Propagat.*, AP-34 (1986): 21–29.
48. G. L. James, "Propagation and Radiation from Partly Filled Elliptical Waveguide," *IEE Proc.*, vol. 136, pt. H (1989): 195–201.
49. P. R. Clark and G. L. James, "Ultra-wideband Hybrid-mode Feeds," *Electron. Lett.*, 31 (1995): 1968–1969.
50. P. R. Clark and G. L. James, "Analysis of Hybrid-mode Feed Horns with Simulated Dielectric Material," *IEEE AP-S International Symposium Digest*, vol. 2 (1994): 772–775.
51. E. Lier and A. A. Kishk, "A New Class of Dielectric-loaded Hybrid-mode Horn Antennas with Selective Gain: Design and Analysis by Single Mode Model and Method of Moments," *IEEE Trans. Antennas & Propagat.*, AP-53 (January 2005): 125–138.
52. G. F. Koch and H. Scheffer, "Coaxial Radiator as Feed for Low Noise Paraboloidal Antennas," *Nachrichtentech. Z.*, vol. 22 (1969): 166–173.
53. M. E. J. Jeuken, M. H. Knoben, and K. J. Wellington, "A Dual Frequency, Dual Polarized Feed for Radio-astronomical Applications," *Nachrichtentech. Z.*, vol. 25 (1972): 374–376.
54. T. S. Bird, G. L. James, and S. J. Skinner, "Input Mismatch of TE₁₁ Mode Coaxial Waveguide Feeds," *IEEE Trans. Antennas Propagat.*, AP-34 (1986): 1030–1033.
55. T. S. Bird, "TE₁₁ Mode Excitation of Flanged Circular Coaxial Waveguides with an Extended Center Conductor," *IEEE Trans. Antennas Propagat.*, AP-35 (1987): 1358–1366.
56. G. L. James, P. R. Clark, G. R. Graves, and T. S. Bird, "The S/X-band Feed for the Tasmanian Earth Resources Satellite Station (TERSS)," *J. Electr. Electron. Eng. Aust.*, vol. 15 (1995): 307–314.
57. R. I. Henderson and P. J. Richards, "Compact Circularly-polarised Coaxial Feed," *IEE International Conference on Antennas and Propagation (ICAP)*, (1995): 327–330.
58. T. S. Bird, "Coaxial Feed Array for a Short Focal-Length Reflector," *IEEE AP-S International Symposium Digest*, vol. 3 (1997): 1618–1621.
59. P.-S. Kildal, "A Small Dipole Fed Resonant Reflector Antenna with High Efficiency, Low Cross-polarization, and Low Sidelobes," *IEEE Trans. Antennas Propagat.*, AP-33 (1985): 1386–1391.

60. G. T. Poulton and T. S. Bird, "Improved Rear-radiating Cup Feeds," *IEEE Antennas Propagation Symposium*, vol. 1, Philadelphia, PA (June 9–13, 1986): 79–82.
61. R. Schwerdtfeger, "A Coaxial Dual Mode Feed System," *IEEE Antennas Propagation Symposium*, vol. 1, Seattle, WA (June 18–22, 1979): 286–289.
62. T. S. Bird, "A Ring-slot Flange for Rear-radiating Cup Feeds," *IEEE Antennas Propagation Symposium*, vol. II, Blacksburg (1987): 672–675.
63. P.-S. Kildal, "The Hat Feed: A Dual-mode Rear Radiating Waveguide Antenna Having Low Cross-Polarization," *IEEE Trans. Antennas Propagat.*, AP-35 (1987): 1010–1016.
64. P.-S. Kildal and J. Yang, "FDTD Optimization of the Bandwidth of the Hat Feed for mm-wave Reflector Antennas," *IEEE AP-S International Symposium*, vol. 3 (1997): 1638–1641.
65. A. D. Olver and J. Xiang, "Design of Profiled Corrugated Horns," *IEEE Trans. Antennas Propagat.*, vol. AP-36 (July 1988): 936–940.
66. C. del Rio, R. Gonzalo, and M. Sorolla, "High Purity Gaussian Beam Excitation by Optimal Horn Antenna," *Proceedings of ISAP* (1996): 1133–1136.
67. G. G. Gentili, E. Martini, R. Nesti, and G. Pelosi, "Performance Analysis of Dual Profile Corrugated Circular Waveguide Horns for Radioastronomy Applications," *IEE Proc.-Microw. Ant. and Prop.*, vol. 148 (2001): 119–122.
68. C. Granet and T. S. Bird, "Optimization of Corrugated Horn Radiation Patterns via a Spline-profile," *ANTEM*, Montréal, Canada (July 27–29, 2002): 307–310.
69. H. Deguchi, M. Tsuji, and H. Shigesawa, "Compact Low-cross-polarization Horn Antennas with Serpentine-shaped Taper," *IEEE Trans. Antennas Propagat.*, AP-52 (2004): 2510–2516.
70. C. Granet, G. L. James, R. Bolton, and G. Moorey, "A Smooth-walled Spline-profile Horn as an Alternative to the Corrugated Horn for Wide Band Millimeter-wave Applications," *IEEE Trans. Antennas Propagat.*, AP-52 (2004): 848–854.
71. G. L. James, "Cross-polarization Performance of Flanged Cylindrical and Conical Waveguides," *Proc. IREE (Aust.)*, vol. 40 (1979): 180–184.
72. B. McA. Thomas, K. J. Greene, and G. L. James, "A Wide-band Prime-focus Horn for Low-noise Receiver Applications," *IEEE Trans. Antennas Propagat.*, vol. 38 (November, 1990): 1898–1900.
73. Z. Ying, A. A. Kishk, and P.-S. Kildal, "Broadband Compact Horn Feed for Prime-focus Reflectors," *Electron. Lett.*, vol. 31 (1995): 1114–1115.
74. T. Scharfen, J. Nellen, and F. van den Bogaart, "Longitudinally Slotted Conical Horn Antenna with Small Flare Angle," *IEE Proc.*, vol. 128, pt. H (June 1981): 117–123.
75. M. S. Aly and S. F. Mahmoud, "Propagation and Radiation Behaviour of a Longitudinally Slotted Horn with Dielectric-filled Slots," *IEE Proc.*, vol. 132, pt. H (December 1985): 477–479.
76. S. P. Skobelev and P.-S. Kildal, "Analysis of Conical Quasi-TEM Horn with a Hard Corrugated Section," *IEEE Trans. Antennas Propagat.*, vol. 51 (October 2003): 2723–2731.
77. P. D. Potter, "A New Horn Antenna with Suppressed Sidelobes and Equal Beamwidths," *Microwave J.*, vol. 6 (June 1963): 71–78.
78. B. MacA. Thomas, "Prime-focus One- and Two-hybrid-mode Feeds," *Electron. Lett.*, vol. 6 (1970): 460–461.
79. T. B. Vu and Q. H. Vu, "Optimum Feed for Large Radiotelescopes: Experimental Results," *Electron. Lett.*, vol. 6 (March 19, 1970): 159–160.
80. R. F. Thomas and D. A. Bathkar, "A Dual Hybrid Mode Feed Horn for DSN Antenna Performance Enhancement," *JPL DSN Progress Report 42-22* (1974): 101–108.
81. A. W. Rudge and N. A. Adatia, "New Class of Primary-feed Antennas for Use with Offset Parabolic-Reflector Antennas," *Electron. Lett.*, vol. 11 (1975): 597–599.
82. A. W. Love, "Two Hybrid Mode, Earth Coverage Horns for GPS," *Dig. Int. Symp. Antennas Propagat.*, Vancouver (June 1984): 575–578.
83. C. Granet, T. S. Bird, and G. L. James, "Compact Multimode Horn with Low Sidelobes for Global Earth Coverage," *IEEE Trans. Antennas & Propagat.*, AP-48 (July 2000): 1125–1133.
84. H. Wolf, "The Scrimphorn, a New Compact Multimode Horn for Array Application," *Journées Internationales de Nice sur les Antennes – JINA* (1988): 446–449.

85. M. Takabayashi, H. Deguchi, S. Makino, and T. Katagi, "A Design Method of Double Bands Dual Mode Conical Horn Antenna by Using Generalized Transmission Equation," *Antennas and Propagation Society International Symposium*, vol. 2 (1996): 922–925.
86. H. Lai, R. Franks, D. Kong, D. Kuck, and T. Gackstetter, "A Broadband High Efficiency Quad-ridged Horn," *IEEE Antennas Propagation Symposium, Blacksburg*, vol. 1 (1987): 676–679.
87. G. L. James, P. R. Clark, and K. J. Greene, "Diplexing Feed Assemblies for Application to Dual-reflector Antennas," *IEEE Trans. Antennas Propagat.*, AP-51 (2003): 1024–1029.
88. H. Z. Zhang and G. L. James, "Characteristics of Quad-ridged Coaxial Waveguides for Dual-band Horn Applications," *IEE Proc.—Microw. Antennas Propag.*, vol. 145: 223–228.
89. A. W. Love, "The Diagonal Horn Antenna," *Microwave J.*, vol. 5 (March 1962): 117–122.
90. J. F. Johansson and N. D. Whyborn, "The Diagonal Horn as a Sub-millimeter Wave Antenna," *IEEE Trans. Microw. Theory Tech.*, vol. 40: 795–800.
91. K. E. Müller, "Untersuchung des Strahlungsverhaltens Elliptischer Hohlleiter Sowie der Möglichkeit zur Erzeugung eines Zirkular Polarisierten Strahlungsfeldes," *Hochfrequenztechn. u. Elektroak.*, vol. 69 (1960): 140–151.
92. M. E. J. Jeuken and L. F. G. Thurlings, "The Corrugated Elliptical Horn Antenna," *IEEE Antennas Propagation Symposium* (1975): 9–12.
93. R. F. E. Guy and R. W. Ashton, "Cross-polar Performance of an Elliptical Corrugated Horn Antenna," *Electron. Lett.*, vol. 15 (1979): 400–402.
94. V. J. Vokurka, "Elliptical Corrugated Horn for Broadcasting Satellite Antennas," *Electron. Lett.*, vol. 15 (1979): 652–654.
95. F. Intoppa, "Breitbandiges Elliptisches Erregensystem," *Antennen*, (1982): 111–114.
96. G. L. James, "Propagation and Radiation from Partly Filled Elliptical Waveguide," *IEE Proc.*, vol. 136, pt. H (June 1989): 195–201.
97. E. Lier, "Broad-band Elliptical Beamshape Horns with Low Cross Polarization," *IEEE Trans. Antennas Propagat.*, vol. AP-38 (June 1990): 800–805.
98. E. Lier, Y. Rahmat-Samii, and S. R. Rengarajan, "Application of Rectangular and Elliptical Dielectric Feed Horns to Elliptical Reflector Antennas," *IEEE Trans. Antennas Propagat.*, AP-39 (1991): 1592–1597.
99. S. Amari and J. Bornemann, "A Study of the Symmetry of Field Patterns of Elliptic Horns Fed by Elliptic Apertures," *IEEE Antennas Propagation Symposium* (1996): 1992–1995.
100. F. R. Cooray, "Analysis of Radiation from a Horn with a Superquadric Aperture," *IEEE Trans. Antenna Propagat.*, vol. 53 (October 2005): 3255–3261.
101. K. Nagasawa and I. Matsuzuka, "Radiation Field of Biconical Horn Antenna with Oblique Edges," *IEEE Antennas and Propagation Society Int. Symp.*, vol. 1 (May 7–11, 1990): 527–530.
102. D. G. Kiely, *Dielectric Aerials* (London: Methuen and Co. Ltd., 1953).
103. R. Chatterjee, *Dielectric and Dielectric Loaded Antennas* (New York: John Wiley & Sons Inc., 1985).
104. C. Salema, C. Fernandes, and R. K. Jha, *Solid Dielectric Horn Antennas* (Boston: Artech House, 1998).
105. N. Brooking, P. J. B. Claricoats, and A. D. Olver, "Radiation Pattern of Pyramidal Dielectric Waveguides," *Electron. Lett.*, vol. 10 (1974): 33–34.
106. R. K. Jha, D. K. Misra, and L. Jha, "Comparative Study of Dielectric and Metallic Horn Antennas," *IEEE Antennas & Propagation Symposium* (July 13–18, 1975): 16–18.
107. S. K. Palit, "Design of Hollow Dielectric Pyramidal Horn Antennas," *IEEE Antennas & Propagation Symposium*, vol. 2 (July 13–18, 1997): 1086–1089.
108. N. V. Venkatarayalu, C.-C. Chen, F. L. Teixeira, and R. Lee, "Numerical Modelling of Ultrawide and Dielectric Horn Antennas Using FDTD," *IEEE Trans. Antennas & Propagat.*, AP-52 (May 2004): 1318–1323.
109. W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas* (New York: Academic Press, 1970).
110. L. A. Weinstein, *The Theory of Diffraction and the Factorization Method* (Boulder CO: The Golem Press, 1969).

111. G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves*, 3rd Ed. (London: Peter Peregrinus Ltd., 1986).
112. P. M. Russo, R. C. Rudduck, and L. Peters, "A Method for Computing E-plane Patterns of Horn Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-13 (March 1965): 219–224.
113. J. S. Yu, R. C. Rudduck, and L. Peters, "Comprehensive Analysis for E-plane of Horn Antennas by Edge Diffraction Theory," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (March 1966): 138–149.
114. C. A. Mentzer, L. Peters, and R. C. Rudduck, "Slope Diffraction and Its Application to Horns," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (March 1975): 153–159.
115. S. G. Hay, F. R. Cooray, and T. S. Bird, "Accurate Modelling of Edge Diffraction in Arrays of Circular and Rectangular Horns," *Journées Internationales de Nice sur les Antennes - JINA* (1996): 645–648.
116. Y. L. Luke, *The Special Functions and Their Approximations*, vol. 2 (New York: Academic Press, 1969): 422–435.
117. S. Schelkunoff, "On Diffraction and Radiation of Electromagnetic Waves," *Phys. Rev.*, vol. 56 (August 15, 1939): 308–316.
118. E. V. Jull, "Finite Range Gain of Sectoral and Pyramidal Horns," *Electron. Lett.*, vol. 6 (October 15, 1970): 680–681.
119. E. V. Jull and L. E. Allen, "Gain of an E-plane Sectoral Horn—A Failure of the Kirchhoff Method and a New Proposal," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (March 1974): 221–226.
120. E. V. Jull, "On the Behavior of Electromagnetic Horns," *IEEE Proc.*, vol. 56 (January 1968): 106–108.
121. E. V. Jull, "Errors in the Predicted Gain of Pyramidal Horns," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (January 1973): 22–31.
122. L. Lewin, *Advanced Theory of Waveguides* (London: Iliffe & Sons, Ltd., 1951).
123. E. V. Jull, "Reflection from the Aperture of a Long E-plane Sectoral Horn," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (January 1972): 62–68.
124. Computer Simulation Technology, CST Microwave Studio, www.cst.de/Content/Products/MWS/Overview.aspx.
125. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (New York: Dover, 1965).
126. K. K. Agarwal and E. R. Nagelberg, "Phase Characteristics of a Circularly Symmetric Dual-mode Transducer," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-18 (December 1970): 69–71.
127. W. J. English, "The Circular Waveguide Step-discontinuity Mode Transducer," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-21 (October 1973): 633–636.
128. T. Satoh, "Dielectric Loaded Horn Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (March 1972): 199–201.
129. K. Tomiyasu, "Conversion of TE_{11} Mode by a Large Diameter Conical Junction," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17 (May 1969): 277–279.
130. R. H. Turrin, "Dual Mode Small Aperture Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-15 (March 1967): 307–308.
131. P.-S. Kildal, E. Lier, and J. A. Aas, "Artificially Soft and Hard Surfaces in Electromagnetics and Their Applications," *IEEE Symp. Antennas Propagat.*, Syracuse (June 1988): 832–835.
132. E. Lier and P.-S. Kildal, "Soft and Hard Horn Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-36 (August 1988): 1152–1157.
133. P.-S. Kildal, "Artificially Soft and Hard Surfaces in Electromagnetics," *IEEE Trans. Antennas Propagat.*, vol. AP-38 (October 1990): 1537–1544.
134. P.-S. Kildal, A. H. Kishk, and S. Maci, "Special Issue on Artificial Magnetic Conductors, Soft/hard Surfaces, and Other Complex Surfaces," *IEEE Trans. Antennas Propagat.*, vol. 53 (January 2005): 2–7.
135. T. S. Chu and W. E. Legg, "Gain of Corrugated Conical Horns," *IEEE Trans. Antennas Propagat.*, vol. AP-30 (July 1982): 698–703.

136. G. L. James, private communication, October 1991.
137. P. F. Goldsmith, *Quasioptical Systems—Gaussian Beam Quasioptical Propagation and Applications* (Piscataway NJ: IEEE Press, 1998).
138. C. Aubry and D. Bitter, "Radiation Pattern of a Corrugated Conical Horn in Terms of Laguerre-Gaussian Functions," *Electron. Lett.*, vol. 11 (1975): 154–156.
139. J. Teniente, D. Goni, R. Gonzalo, and C. del-Rio, "Choked Gaussian Antenna: Extremely Low Sidelobe Compact Antenna Design," *IEEE Antennas and Wireless Propagation Letters*, vol. 1 (2002): 200–202.
140. N. W. McLachlan, *Theory and Application of Mathieu Functions* (New York: Dover, 1964).
141. "FORTRAN Routines for Computation of Special Functions," <http://jin.ece.uiuc.edu/routines/routines.html>.
142. Wolfram Research, "Mathematica," <http://www.wolfram.com/>.
143. E. Lier, "An Elliptical Corrugated Horn Model for Radiation Pattern Predictions," *IEEE Antennas & Propagation Symposium*, vol. 1 (June 18–23, 1995): 156–159.
144. V. Venkatesan and K. T. Selvan, "Rigorous Gain Measurement on Wide-band Ridge Horn," *IEEE Trans. Electromag. Compatibility*, vol. 48 (August 2006): 592–594.
145. E. Lier, "Bandwidth of a Square Hard Horn," *IEE Proc.*, vol. 135, pt. H (1988): 275–278.
146. A. W. Love, "Hybrid Mode Horn Antennas," *Electromagnetics*, vol. 6 (1986): 315–332.
147. C. M. Knop, Y. B. Cheng, and E. L. Ostertag, "On the Fields in a Conical Horn Having an Arbitrary Wall Impedance," *IEEE Trans. Antennas Propagat.*, vol. 34 (September 1986): 1092–1098.
148. C. S. Lee, S. W. Lee, and D. W. Justice, "A Simple Circular Polarized Antenna: Circular Waveguide Horn Coated with Lossy Magnetic Material," *IEEE Trans. Antennas Propagat.*, vol. AP-36 (February 1988): 297–300.
149. J. J. H. Wang and V. K. Tripp, "Design and Performance of the Magnetic Hybrid Mode Horn," *IEEE Trans. Antennas Propagat.*, vol. AP-37 (November 1989): 1407–1414.
150. J. J. Lee, "Numerical Methods Make Lens Antennas Practical," *Microwaves* (September 1982): 81–84.
151. E. L. Holzman, "A Highly Compact 60GHz Lens Corrected Conical Horn Antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 3 (2004): 280–282.

Chapter 15

Reflector Antennas

Yahya Rahmat-Samii

University of California, Los Angeles

CONTENTS

15.1 INTRODUCTION	15-2
15.2 DIFFRACTION ANALYSIS TECHNIQUES FOR REFLECTOR ANTENNAS	15-3
15.3 CONVENTIONAL REFLECTOR ANTENNAS	15-10
15.4 OFFSET (SYMMETRIC) PARABOLOIDAL REFLECTORS	15-16
15.5 REFLECTORS OTHER THAN PLANE OR PARABOLIC	15-32
15.6 PASSIVE REFLECTORS	15-40
15.7 LARGE-APERTURE REFLECTOR ANTENNAS	15-47
15.8 DIFFRACTION SHAPING	15-57

The original chapter was written by Kenneth S. Kelleher and Geoffrey Hyde.

15.1 INTRODUCTION

Demand for reflector antennas for use in radars, radio astronomy, microwave communications, satellite communications and tracking, remote sensing, and the like has resulted in the development both of sophisticated reflector antenna configurations and of analytical and experimental design techniques. Reflector antennas may take varied configurations, some of the most popular ones being plane, corner, singly and doubly curved, and single and multiple reflectors. The actual configurations used depend heavily on the nature of applications and the resulting radiation patterns such as pencil beam and contour beam, as well as multifrequency and polarizations, etc.

Searching the literature, we see that we can classify reflector antennas in a variety of ways. Reflector antennas can be identified according to *pattern* type, *reflector surface* type, and *feed* type. Pencil-beam reflectors are perhaps the most popular ones and are commonly used in point-to-point microwave communications and telemetry because their patterns yield the maximum boresight gain, and typically their beam directions are fixed at the time of antenna installation. In some applications, such as satellite communication systems, the uplink beam of these pencil-beam reflectors may be either fully steerable by reflector movements, or capable of limited steering. Recent generations of satellite reflectors have produced other popular types of pattern classifications: contour (shaped) beams and multiple beams. These applications require reflectors with improved off-axis beam characteristics, which result in more sophisticated configurations. There are many microwave communication antennas operating with one sense of polarization at a given frequency and require only reasonable discrimination between orthogonal polarizations. However, the current generation of microwave communication and remote sensing antennas operate with dual polarizations at the same frequency to enhance their so-called frequency reuse capabilities or radar polarimetric properties. These requirements necessitate improved polarization performance and could be used as factors in pattern classification.

Performance of reflector antennas cannot be properly examined without knowledge of their feed configurations. Horn or waveguide feeds operating in a single pure mode have been the most popular feeds for reflectors. However, to meet radio astronomy, earth station, and satellite antenna requirements, considerable effort has been focused on the development of new feeds to efficiently illuminate either the main reflector or the subreflector of the antenna. For example, hybrid-mode feeds (combining TE and TM fields) are used to match efficiently the feed distribution with the desired focal field distribution of the reflector, which also results in an ideal feed for reducing cross polarization. Among the hybrid-mode feeds are feeds with corrugated metal walls. Another important consideration is the complexity of the feed system in terms of the number of feed elements used, for instance, to create contour or multiple beams. The proximity of feed elements necessitates an understanding of mutual coupling and the overcoming of difficulties in designing an acceptable beamforming network. Even more recently microstrip patch antennas have been considered as array feeds for reflector antennas. Impulse radiating reflector antennas and compact range reflector antennas require feeds of very wide bandwidths. Due to page limitations, this chapter does not focus on various feeds and their designs. The reader may need to consult other chapters of this handbook or other pertinent references.

The objective of this chapter is to enhance the material in the earlier version of this chapter written by K. Kelleher and G. Hyde¹ and to use updated material and newly generated graphs based on Rahmat-Samii's book chapter² on reflector antennas and recent publications to provide a summary of the performances of many representative reflector antenna configurations as just discussed. Numerous design charts of offset reflector configurations,

which are used extensively in today's reflector antenna designs, are provided. There are a few recently published books, review papers, and collections of papers that the reader is strongly advised to study.¹⁻¹¹ These references provide detailed mathematical, simulation, and measurement results.

15.2 DIFFRACTION ANALYSIS TECHNIQUES FOR REFLECTOR ANTENNAS

The maturity and accuracy of analytical/numerical techniques have allowed reflector antenna engineers to extensively utilize commercial or customized computer programs in the performance evaluation and design of reflector antennas. There are ample successful checks of the results of these techniques against measured data. This section describes the basics of some of the current and popular techniques. As a baseline, dual-offset reflector antennas are used for the description of these techniques.

Figure 15-1 shows the geometry of a dual-offset reflector with a feed arbitrarily positioned. This is a rather general configuration, and all other cases, such as single reflectors, symmetric configurations, Cassegrains, etc., can be regarded as special cases. Reflectors with more than one subreflector (beam waveguide systems) may also be handled by replacing the subreflector of Figure 15-1 with the appropriate beam waveguide systems. To facilitate the discussion, three coordinate systems are erected to define the main reflector, the subreflector, and the feed (or array of feeds) position. The position and field vectors of these coordinate systems can be interrelated using the Eulerian angles construction.^{12,13} Using appropriate coordinate transformations, the fields of a feed can be expressed in feed coordinates (x_F, y_F, z_F), then transformed into subreflector coordinates (x_S, y_S, z_S) to determine the scattered field from the subreflector, and then transformed again into main reflector coordinates (x_M, y_M, z_M) to finally obtain the radiated field of the main reflector.

Among many analytical/numerical diffraction analysis techniques, we can refer to physical optics (PO), aperture field (AF), geometrical optics (GO), geometrical theory of diffraction (GTD), method of moments (MOM), Gaussian beam method, or any combination of these. Clearly, advantages and disadvantages appear when comparing these techniques. Comparisons can be made based on the particular reflector configuration,

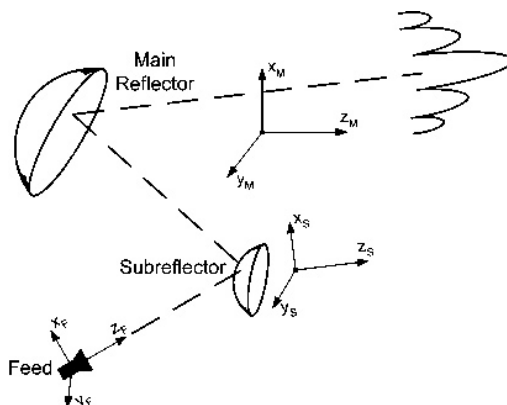


FIGURE 15-1 Geometry of a dual-offset reflector antenna with various coordinate systems

far-field pattern domain, polarization, computation time, accuracy, etc. MOM gives the most accurate result, but it is impossible to use it economically for reflectors larger than several wavelengths. The aperture field method is not very accurate for offset configuration with displaced feeds when the edge diffracted rays are not included in the construction of the aperture fields. The GTD method is not easily applied in the caustics regions of pencil beams. The PO method can take an excessive amount of computation time for large reflectors, in particular, in the dual-reflector configuration, etc. Many advances have been reported to improve both the accuracy and the computation time. One of the most appealing techniques has been the application of GTD for the subreflector and physical optics for the main reflector, in conjunction with efficient expansions, such as Jacobi-Bessel, sampling theorem, etc. A summary of some of these techniques that utilize some of the Rahmat-Samii's materials² is provided next.

Physical Optics (PO) Technique

The mathematical formulation of physical optics, or PO, is based on the assumption that the induced current on a perfect electric conductor (PEC) reflector surface is given by

$$\mathbf{J} = \begin{cases} 2\hat{\mathbf{n}} \times \mathbf{H}^i & \text{illuminated region} \\ 0 & \text{otherwise} \end{cases}$$

where $\hat{\mathbf{n}}$ is the unit normal to the surface and \mathbf{H}^i is the incident magnetic field. This incident field may emanate directly from the source or be scattered from the subreflector. Although PO current is an approximation for the true current on the reflector surface (obtainable using MOM), it nevertheless gives very accurate results for far fields of reflectors as small as approximately 5 wavelengths in diameter. The PO current displays certain errors, particularly near the edge of the reflector, which can be augmented by the incorporation of the fringe current. Unfortunately, the incorporation of the fringe current is not in general an easy task, although it may be required to accurately determine the far fields for a very wide range of observation angles, typically near 90° off boresight. The radiated \mathbf{E} field can be constructed using

$$\mathbf{E}(\mathbf{r}) = -\frac{j}{\omega\epsilon} \int_{\Sigma} [(\mathbf{J} \cdot \nabla) \nabla + k^2 \mathbf{J}] g(R) dS' \quad (15-1)$$

which is applicable for both the near- and far-field zones.^{5,14} The preceding expression can be easily derived using the concept of vector potentials. In Eq. 15-1, \mathbf{r}' , \mathbf{r} , and \mathbf{R} are described in Figure 15-2, $k = \omega\sqrt{\mu\epsilon}$, and

$$\begin{aligned} g(R) &= \frac{e^{-jkR}}{4\pi R} \\ \nabla g(R) &= \left(jk + \frac{1}{R} \right) g(R) \hat{\mathbf{R}} \\ (\mathbf{J} \cdot \nabla) \nabla g(R) &= \left[-k^2 (\mathbf{J} \cdot \hat{\mathbf{R}}) \hat{\mathbf{R}} + \frac{3}{R} \left(jk + \frac{1}{R} \right) \right. \\ &\quad \left. \times (\mathbf{J} \cdot \hat{\mathbf{R}}) \hat{\mathbf{R}} - \frac{1}{R} \left(jk + \frac{1}{R} \right) \mathbf{J} \right] g(R) \end{aligned} \quad (15-2)$$

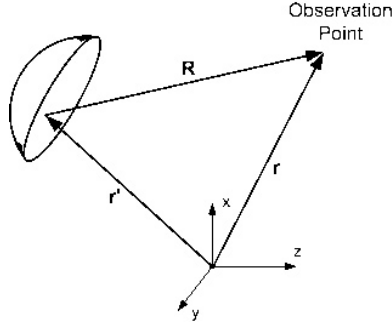


FIGURE 15-2 Reflector geometry for constructing the physical optics (PO) formulation

Further simplification of Eq. 15-1 allows us to obtain the following well-known expression for the radiated field in the far-field zone:

$$\mathbf{E}(\mathbf{r}) = -\frac{jk^2}{\omega\epsilon} g(r)(\bar{\mathbf{I}} - \hat{\mathbf{r}}\hat{\mathbf{r}}) \cdot \mathbf{T}(\theta, \phi) \quad (15-3)$$

$$\mathbf{T}(\theta, \phi) = \int_{\Sigma} \mathbf{J} e^{jk\hat{\mathbf{r}} \cdot \mathbf{r}'} d\Sigma' \quad (15-4)$$

In deriving the preceding expression the standard far-field approximations are used:

$$R = |\mathbf{r} - \mathbf{r}'| \cong r - \hat{\mathbf{r}} \cdot \mathbf{r}'$$

A similar but slightly more complicated expression can also be developed for the field in the Fresnel zone.

The curved surface integration in Eq. 15-4 over Σ can be recast in terms of integration variables defined in a projected planar aperture by properly using the surface projection transformation

$$\mathbf{T}(\theta, \phi) = \int_{\sigma} \mathbf{J}_{\text{eff}} e^{jk\hat{\mathbf{r}} \cdot \mathbf{r}'} dS' \quad (15-5)$$

where

$$\mathbf{J}_{\text{eff}} = \mathbf{J} \sqrt{1 + \left(\frac{\partial f}{\partial x'}\right)^2 + \left(\frac{\partial f}{\partial y'}\right)^2}$$

In Eq. 15-5 σ is the projected aperture in the xy plane, and $z' = f(x', y')$ defines the description of the reflector surface.

The physical optics radiation integral in Eq. 15-5 may be evaluated in many different ways, and its efficient evaluation has been a challenging problem. A successful numerical method has been used based on the small patch constant phase technique,¹⁵ which can be applied to both the subreflector and main reflector geometries. Efficient numerical methods, which are particularly applicable to the main reflector with circular and elliptical

projected apertures, have been developed in Rahmat-Samii and Galindo-Israel¹⁶ and in Galindo-Israel and Mittra.¹⁷ One of these methods, which is based on the Jacobi-Bessel expansion, is particularly effective when it is used to evaluate Eq. 15-5. It can be shown that, based on this expansion, the radiation integral can be expressed as¹⁶

$$T \propto \sum_{p=0}^{P \rightarrow \infty} \frac{1}{p} (jk)^p (\cos \theta - \cos \theta_B)^p \sum_{n=0}^{N \rightarrow \infty} \sum_{m=0}^{M \rightarrow \infty} J^n [{}_p C_{nm} \cos n\phi + {}_p D_{nm} \sin n\phi] \\ \times \sqrt{2(n+2m+1)} \left[\frac{J_{n+2m+1}(kaB)}{kab} \right] \quad (15-6)$$

In the previous equation, a is the radius of the geometrically projected circular aperture, J is the Bessel function, and

$$B = \sqrt{(\sin \theta \sin \phi - \sin \theta_B \sin \phi_B)^2 + (\sin \theta \cos \phi - \sin \theta_B \cos \phi_B)^2} \\ \phi = \tan^{-1} \left(\frac{\sin \theta \sin \phi - \sin \theta_B \sin \phi_B}{\sin \theta \cos \phi - \sin \theta_B \cos \phi_B} \right) \quad (15-7)$$

where (θ_B, ϕ_B) is the direction of the anticipated beam maximum (or its vicinity). It should be noted that when $B = 0$, i.e., in the boresight direction, only the first term of the series contributes. This first term is the Airy disk function, a typical far-field pattern of a uniform circular aperture. Another important feature of Eq. 15-6 is that the coefficients C and D do not depend on the observation points. These coefficients take the following form:

$$\begin{Bmatrix} {}_p C_{nm} \\ {}_p D_{nm} \end{Bmatrix} = \frac{\epsilon_n}{2\pi} \int_0^{2\pi} \int_0^1 \mathbf{Q}_p \begin{Bmatrix} \cos n\phi' \\ \sin n\phi' \end{Bmatrix} F_m^n(s') s' d\phi' ds' \quad (15-8)$$

where ϵ_n is the Neumann constant ($\epsilon_n = 1$ for $n = 1$, and $\epsilon_n = 2$ for otherwise), F_m^n is the modified Jacobi polynomial,¹⁸ and \mathbf{Q}_p is directly related to the physical optics induced current. The exact mathematical expression of \mathbf{Q}_p can be found in Rahmat-Samii and Galindo-Israel¹⁶ and Rahmat-Samii et al.¹⁸

Upon the observation that the far-field pattern is related to the Fourier transform of the aperture field, an efficient representation of the far field can be obtained by applying the sampling theorem, which allows you to express the far fields from a knowledge of fields at a limited number of sampling points,¹⁹ namely,

$$\mathbf{E} = \sum \sum \mathbf{E}_{nm} \frac{\sin(u - n\pi)}{u - n\pi} \frac{\sin(v - m\pi)}{v - m\pi} \quad (15-9)$$

where $u = ka' \sin \theta \cos \phi$ and $v = ka' \sin \theta \sin \phi$ (a' is the radius of an extended aperture¹⁹ in front of the reflector) and the \mathbf{E}_{nm} are the fields at the sampling points dictated by the Witterker-Shannon theorem with spacing $\Delta u \leq \pi/a'$ and $\Delta v \leq \pi/a'$. This far-field representation is particularly useful when you are evaluating the fields at many observation points. A generalization of this procedure can also be obtained using nonuniform sampling techniques.²⁰

Geometrical Theory of Diffraction (GTD) Technique

Because the fundamentals of the geometrical theory of diffraction (GTD) are discussed in many books and articles, only its application to reflectors is summarized here. The foundation of GTD is based on the concepts of localization and ray picture satisfying Fermat's principle. These rays may be classified as the geometrical and edge-diffracted rays emanating from surface reflection and edge-diffracted points in accordance with Snell's law and Keller's theory.²¹⁻²³ For the reflector shown in Figure 15-3 the total scattered \mathbf{H} field may be expressed as

$$\mathbf{H}' = \theta^i \mathbf{H}^i + \sum_i \theta_i^r \mathbf{H}_i^r + \sum_j \mathbf{H}_j^d \quad (15-10)$$

where the superscripts i , r , and d denote the incident, reflected, and diffracted fields, respectively, and the subscripts i and j refer to the number of surface reflection and edge diffraction specular points, respectively, for a given feed and observation location. Additionally, θ^i and θ^r define the incident and reflected shadow indicators (i.e., $\theta = 0$ if the ray is blocked, and $\theta = 1$ if the ray path is unblocked).^{22,23} The \mathbf{E} field can be constructed in a straightforward manner using the fact that on each ray the \mathbf{E} and \mathbf{H} fields assume a local plane wave relationship, namely,

$$\mathbf{E} = Z \mathbf{H} \times \hat{\mathbf{r}}$$

where Z is the intrinsic impedance of the medium ($Z = 120\pi\Omega$ for the free space) and $\hat{\mathbf{r}}$ is the unit vector in the direction of each ray.

For a single feed with a well-defined phase center, we typically associate rays emanating from the phase center. For more distributed sources (arrays of feeds), you may have to associate many phase centers (one per feed element) and repeat the GTD construction for each of them. For a given feed position and specified observation point (this can be a point on the main reflector), you must first determine the GO reflection and edge-diffracted points (there can be more than one point for each type). There are various techniques for locating these points; however, all of them are based on Fermat's principle, which requires that the optical path be optimal (stationary). Fermat's principle may be invoked by requiring that the functional

$$d = d_{1(3)} + d_{2(4)} \quad (15-11)$$

as defined in Figure 15-3 be stationary with the constraint that Eq. 15-11 gives a solution(s) on the reflector surface and the edge boundary. Upon the determination of reflection and diffraction points from the solution of the preceding nonlinear equations, you can then apply the field construction of GTD to determine both the reflected and diffracted fields. The details of this construction may be found in Lee et al.,²² Rahmat-Samii et al.,²⁴ and other references.

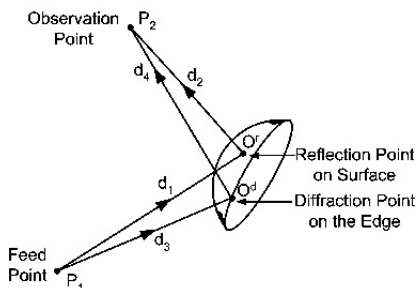


FIGURE 15-3 Reflection and diffraction points for constructing the geometrical theory of diffraction (GTD) formulation along the reflected and diffracted rays

The reflected field transported along the reflected rays can be formulated as

$$\mathbf{H}^r(P_2) = (DF)e^{-jk d_2} \{ \mathbf{H}^i(O^r) - 2[\mathbf{H}^i(O^r) \cdot \hat{\mathbf{n}}] \hat{\mathbf{n}} \} \quad (15-12)$$

which is given in terms of the incident field \mathbf{H}^i at the reflection point O^r , the unit surface normal $\hat{\mathbf{n}}$ at O^r , and a divergence factor (DF). This factor can be expressed as

$$DF = \frac{1}{\sqrt{1 + (d_2/R_1')}} \frac{1}{\sqrt{1 + (d_2/R_2')}} \quad (15-13)$$

where the square roots take positive real, negative imaginary, or zero values (so that DF is positive real, positive imaginary, or infinite). The terms R_1' and R_2' are the radii of principal curvature of the reflected wavefront passing through O^r . Their computation can be found in Lee et al.²²

For each diffraction point O^d in Figure 15-3, there is a contribution to the diffracted field \mathbf{H}^d in Eq. 15-10. The diffracted field can be expressed using the formulation in Lee et al.²²

$$\mathbf{H}^d(P_2) = \frac{1}{2\sqrt{2\pi d_4}} e^{-jk(d_4 + \pi/4)} \frac{1}{\sqrt{1 + (d_4/R_1)}} \frac{1}{\sin \beta} [\hat{\beta} D^h H_\beta^i + \hat{\alpha} D^s H_\alpha^i] \quad (15-14)$$

where

R_1 = the radius of curvature of the diffracted wavefront passing through O^d

β = the angle between the tangent to the edge and \mathbf{d}_4

$\hat{\alpha}, \hat{\beta}$ = the unit vectors of the diffracted-ray coordinates

D^h, D^s = the hard and soft diffraction coefficients

H_α^i, H_β^i = the projections of the \mathbf{H}^i incident field on the ray coordinates at O^d

Note that Keller's standard diffraction coefficients diverge at the incident and reflected shadow boundaries, which can be corrected by applying a uniform diffraction coefficient. There are different uniform expressions for the diffraction coefficients, and the reader may refer to the literature²⁵⁻²⁸ for a detailed comparison. The most popular ones are the uniform theory of diffraction (UTD) and uniform asymptotic theory (UAT).

Aperture Field (AF) Technique

Another popular reflector analysis technique is based on the application of the aperture field, or AF, method. The mathematical construct of this technique requires that first a hypothetical planar aperture is erected in front of the reflector, and then the tangential fields are determined in this aperture using the GO and GTD constructions. It is customary to truncate the aperture to the reflector projected aperture size when dealing with well-focused pencil-beam antennas, and only the GO field construction is used. For cases where the feed is defocused or for fields at wide observation angles, however, you must use larger apertures to properly incorporate the contribution of the edge-diffracted fields. Alternatively you can use an aperture that caps the reflector's rim. Upon the determination of the tangential \mathbf{E} and \mathbf{H} fields in the aperture, the far fields can then be obtained using equivalence theorem.

There are typically three different representations for the construction of far fields that use either the aperture tangential \mathbf{E} or \mathbf{H} fields separately or a combination of them.

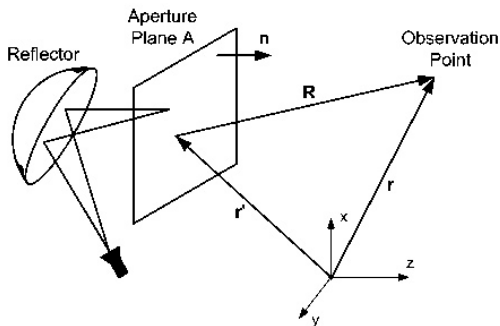


FIGURE 15-4 Reflector and aperture geometry and integration and observation parameters for constructing the aperture field (AF) formulation

These different representations are customarily referred to as different Kirchhoff approximations.⁴ One of the representations that use both the aperture tangential \mathbf{E} and \mathbf{H} fields and results in the far-field radiated tangential \mathbf{E} field can be formulated as

$$\mathbf{E} = \frac{-jk^2}{\omega\epsilon} g(r) \int_A \left\{ \hat{\mathbf{n}} \times \mathbf{H}^A - [(\hat{\mathbf{n}} \times \mathbf{H}^A) \cdot \hat{\mathbf{r}}] \hat{\mathbf{r}} - \frac{1}{Z} [(\hat{\mathbf{n}} \times \mathbf{E}^A) \times \hat{\mathbf{r}}] \right\} e^{jk\hat{\mathbf{r}} \cdot \mathbf{r}'} dA \quad (15-15)$$

where \mathbf{E}^A and \mathbf{H}^A are the \mathbf{E} and \mathbf{H} fields in the aperture with aperture size A and $\hat{\mathbf{n}}$ is the normal to this planar aperture, as shown in Figure 15-4. Variable $g(r)$ has already been defined in Eq. 15-2. For a planar aperture, Eq. 15-15 is a two-dimensional Fourier transform and can be evaluated in many different ways, one of which includes the use of the fast Fourier transform (FFT) algorithm.

Extension of Diffraction Analysis Techniques to Mesh, FSS, and Other Surfaces

There are many applications for which the reflector surface may not be simply a solid PEC surface. For example, mesh reflectors are used in large unfurlable antennas, FSS (frequency selective surfaces) are used as subreflectors for frequency separation of feeds, the reflector surface may be affected by snow, etc. In all these cases, you can modify the PO integral and incorporate the effects of these surfaces.²⁹⁻³³

Near, Fresnel, and Far Fields of a Reflector Antenna

Before closing this section it is worthwhile to present data on the near, Fresnel, and far fields of a typical parabolic reflector antenna. This is useful for developing an insight into the evolution of radiated fields away from the reflector. For the sake of simplicity, only the uniform aperture distribution is considered, and its radiated fields at many different distances from the aperture are constructed by numerically evaluating Eq. 15-1. Results are shown in Figure 15-5, where you can readily observe how the far-field pattern is formed at distance $R > 2D^2/\lambda$. For very low sidelobe reflectors, distances much greater than $2D^2/\lambda$ are required to obtain an accurate description of inner sidelobes. These results are constructed for a symmetric parabolic reflector antenna with diameter of 100 wavelengths and $F/D = 0.4$ with almost no taper illumination.

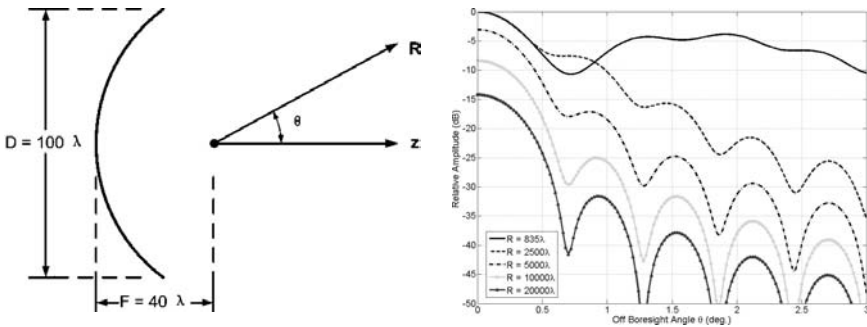


FIGURE 15-5 Radiated field pattern of a reflector antenna as a function of distance R

15.3 CONVENTIONAL REFLECTOR ANTENNAS

The reflector antennas treated in this chapter are those that have been used for decades. The very-high-gain reflectors used in earth stations, which have a more sophisticated design, are treated in the later sections of this chapter. As mentioned earlier, the conventional reflector designs are (1) planar, (2) parabolic-arc, (3) double-curvature, (4) paraboloid, and (5) shielded reflectors. In addition, passive reflectors used in communications and radar are described. More sophisticated large reflectors will be discussed at the end of this chapter.

Corner Reflector³⁴

The corner-reflector antenna is made up of two plane reflector panels and a dipole element (see Figure 15-6). This antenna is useful in obtaining gains of the order of 12 dBi. Higher gains can be obtained by using large reflectors and larger spacing of the dipole to the panel intersection. The gain also depends upon the corner angle. Figure 15-7 shows the gain as a function of dipole spacing for four values of corner angle. Figures 15-8 through 15-11 show the radiation patterns for E and H planes. One of the angles is 180° , and the data therefore include the planar reflector.

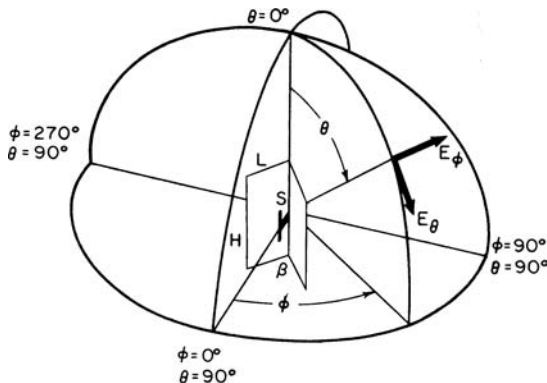


FIGURE 15-6 Corner-reflector antenna and coordinate system

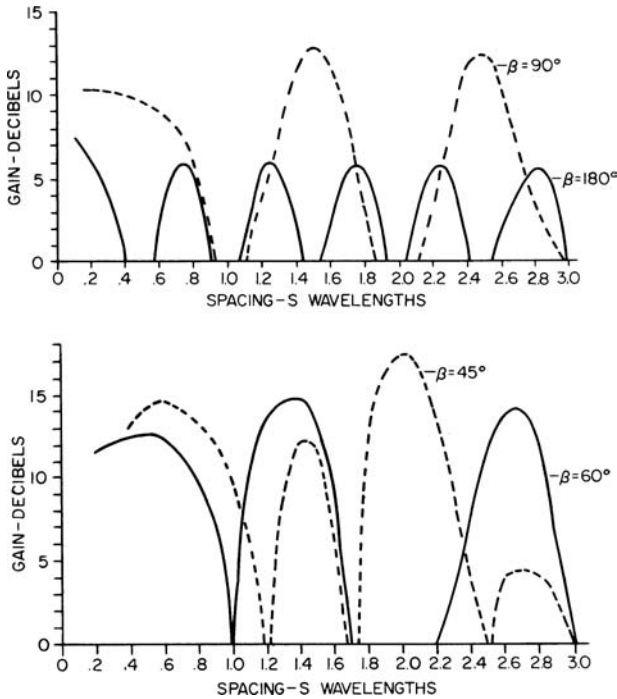


FIGURE 15-7 Gain of corner reflectors

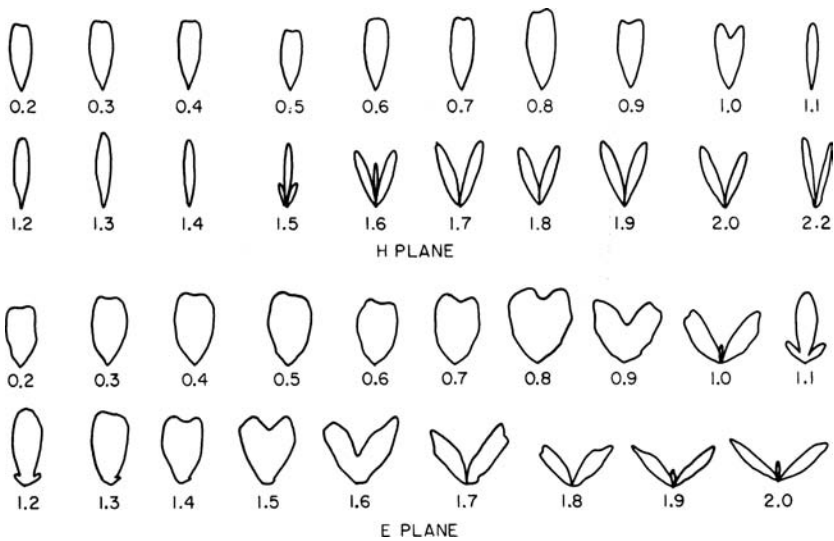


FIGURE 15-8 Patterns of 60° corner, S_λ variable

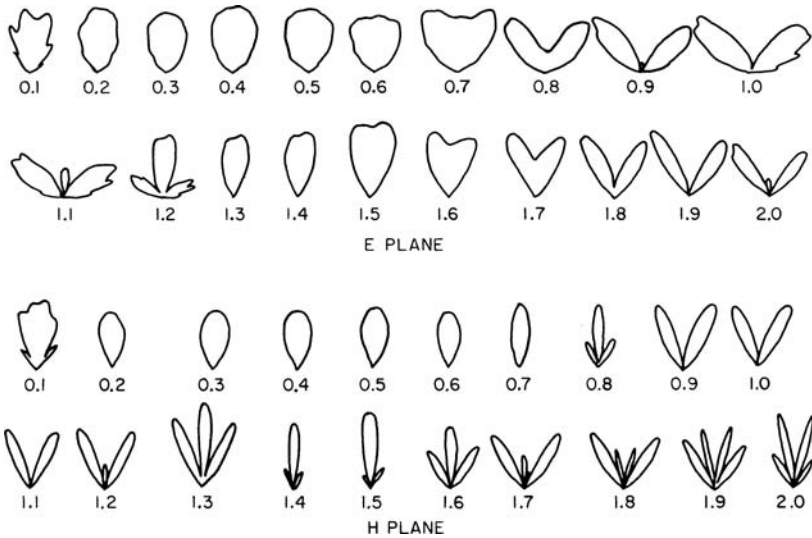


FIGURE 15-9 Patterns of 90° corner, S_λ variable

The size of reflector panels is a compromise between gain and antenna wind load. The wind load is reduced by replacing the panels by parallel rod reflectors with spacing of $\lambda/10$. The rod length should be 0.6λ . The size of the reflector panels must be increased as the dipole spacing is increased. A spacing of $\lambda/3$ requires a panel length of 0.6λ , a spacing of $\lambda/2$ requires λ , etc. The impedance of the corner reflector depends upon the spacing and upon the dipole impedance. Spacing of less than $\lambda/3$ for the 90° corner is not recommended. Larger spacing has less effect on impedance.

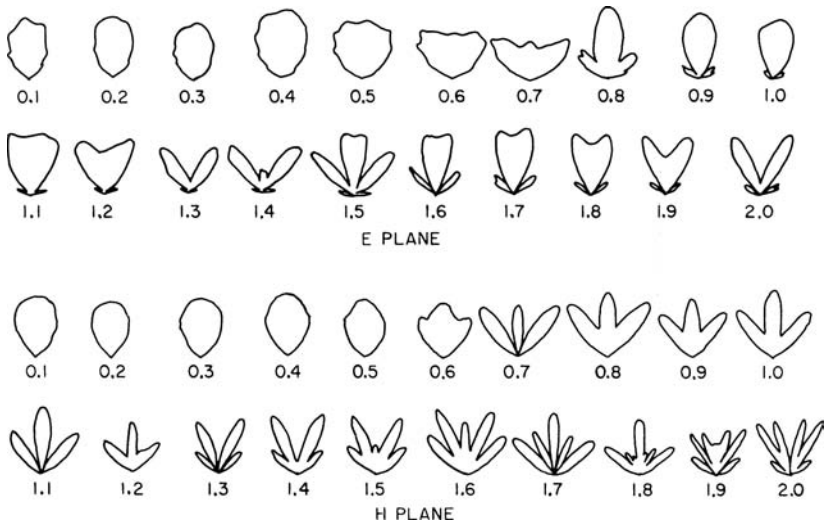


FIGURE 15-10 Patterns of 120° corner, S_λ variable

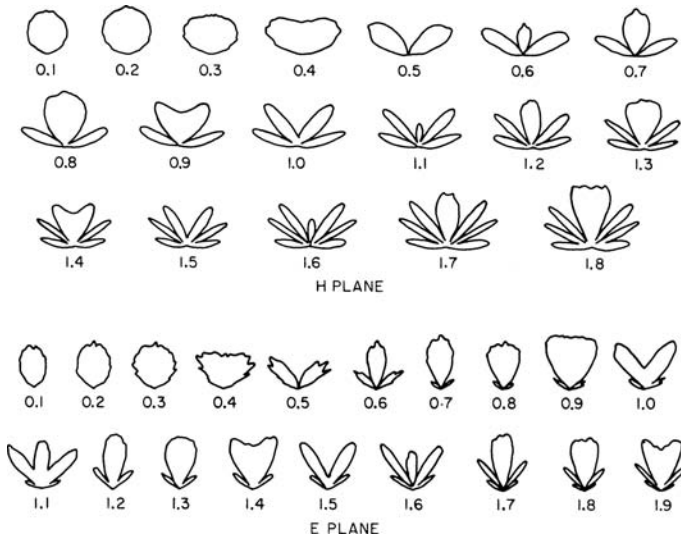


FIGURE 15-11 Patterns of flat-sheet, S_λ variable

An increase in the corner-reflector gain³⁵ is obtained by adding a third panel, which serves as a ground plane for a 0.75λ monopole feed. The long monopole gives a peak signal above the ground plane (about 45°). The panel lengths are 2λ , and the monopole spacing is about one wavelength. As in the conventional corner reflector, the gain depends upon corner angle and spacing: 90° and 0.9λ give 17 dBi, 60° and 1.2λ give 19 dBi, and 45° and 1.6λ give 21 dBi.

An increase in corner-reflector bandwidth³⁶ can be achieved with a broadband dipole feed. An open-sleeve dipole with the center section of coaxial line in the sleeve replaced by slab line gave almost an octave bandwidth with a voltage-standing-wave-ratio (VSWR) of 2.5:1. The gain was 11.5 ± 1 dBi. The match was achieved by a cut-and-try procedure.

Reflectors with Parabolic Arc

The parabolic arc can be used as a generating curve for a number of useful reflector surfaces. When moved so that the focal point travels along a straight line, the curve generates a parabolic cylinder. With a limited motion along this line, the reflector for a pillbox is generated. When the focal point is moved along a circle enclosing the arc, an hourglass surface is generated, and for a circle that does not enclose the arc a parabolic torus is generated. Figure 15-12 shows the basic parameters of these surfaces.

Parabolic Cylinder

The elements of the cylinder are all perpendicular to the plane of the parabolic arc. Because this is a singly curved surface, problems of construction and maintaining tolerances are somewhat easier than those encountered in the case of the paraboloid. The cylinder can be constructed of tubes or slats that are straight-line elements of the surface, or it can use identical parabolic arcs formed of tubes or slats positioned so that the planes of the parabolic arcs are perpendicular to the elements of the cylinder.


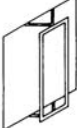
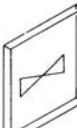


REFLECTOR		DIMENSIONS (WAVELENGTHS)	BANDWIDTH (%)	DIRECTIVE GAIN (dBi)
MICROSTRIP		0.5×0.5 $t = 0.05$	1%	6
SKELETON SLOT		0.5×0.5 $t = 0.25$	80%	9
PLANETENNA		1.0×1.0 $t = 0.1$	30%	11
SHORT BACKFIRE		$D = 2$ $t = 0.5$	10%	15
GRID ARRAY		$D = 5$ $t = 0.05$	15%	22

FIGURE 15-12 Various reflector configurations with parabolic curves

The widest use of the cylinder is with a line-source feed, which may be formed in a number of ways. The line-source aperture is made to coincide with the focal line of the cylinder (the line made by joining the focal points of all arcs). The aperture of the line source can contain energy in phase along the length or have a linear-phase variation along the length. The first produces a cylindrical wavefront, while the linear-phase variation produces a conical wavefront. The cylindrical wave is focused into a plane wave in the direction of the axis of a parabolic arc. The conical wavefront produces a plane wave tilted from the parabolic axis by an angle equal to the half angle of the cone.

The illumination problems of a line-source feed with the cylinder are related to those of the paraboloid. The decrease in energy in the cylindrical or conical wave is proportional to distance rather than to the square of the distance. The sidelobes present a problem in the parabolic cylinder since the relative gain between the line source and the final antenna system is not sufficient to suppress the sidelobes of the source. Care must be given to the design of the aperture of the line source, particularly in control of the *E*-plane sidelobes.

The feed-blocking problem for a line source and parabolic cylinder is serious since the line source presents a large blockage. For this reason, it is usually desirable to use an offset reflector in which the line-source feed structure is removed from the path of the reflected radiation.

The parabolic cylinder can be fed by a point-source feed horn. The horn illumination function is defined by energy that falls off with the square of distance. The reflected wavefront will be a cylinder with elements perpendicular to the elements of the parabolic cylinder. In one plane the point source is focused, and in the other plane the cylinder reproduces the pattern of the feed horn. The result can be a narrow beam in azimuth and a broad-coverage beam in elevation. The best work on such an antenna used a cylinder with its straight-line elements tilted from vertical. Close-in sidelobes were controlled, and the largest sidelobes (of the order of -28 dB) were removed from the main-beam region. This reflector operated at 9.3 GHz and measured 5 ft by 1 ft (1.5 m by 0.3 m), with a focal length of 1.5 ft (4.5 m). The tilt of the reflector was 12° .

Since the horn and cylinder will give a cylindrical wave and the parabolic cylinder focuses the cylindrical wave, a second reflector can be used. Use of a point-source feed and two parabolic cylinders gives a final wavefront identical to that of a paraboloid. If the equation of the first cylinder is $y^2 = 4fx$, with f the focal length, the second cylindrical surface is given by $(x+f)^2 = 4f(z+f)$, and the final rays are all parallel to the z axis. The plane curves of intersection for the cylinders before development are given in Spencer et al.³⁷

The parabolic cylinder can also be used as the reflector in the pillbox, or cheese, antenna. The pillbox is formed by two parallel planes that cut through a parabolic cylinder perpendicular to the cylinder elements. Typically the focal line of the cylinder is positioned in the center of the aperture formed by the open ends of the parallel plates. When a feed is placed at the focal line, it blocks a significant portion of the open region. The blockage causes large sidelobes in the pattern of the pillbox. The feed backlobes are significant, and a portion of the feed's radiation is reflected back into itself to produce a standing wave. An improvement in performance is obtained when a half pillbox is used in the same fashion as the offset parabolic cylinder. The arc of the parabola extends from the vertex to the 90° point. The feed horn is pointed at the 45° point, and although the illumination is asymmetrical, good sidelobes are obtained. The maximum sidelobe of -26 dB was from the spillover past the 90° point of the parabolic arc. A recent spaceborne concept was designed and tested in Rahmat-Samii et al.³⁸

Hourglass Reflector³⁹

The hourglass reflector is generated by rotating a parabolic arc (symmetrical or offset) about a vertical axis that is on the convex side of the arc. The feed system for the hourglass reflector is a circular array. The hourglass serves to give increased gain in the vertical plane. The circular array has the characteristics of the Wullenweber system,⁴⁰ which provides scanning beams, fixed beams, and beams from a hybrid matrix system. The sidelobes in the elevation plane are controlled by controlling the elevation pattern of the array elements. In the horizontal plane the sidelobes are controlled by tapering the energy in the array elements.

Parabolic Torus⁴¹

This reflector is formed by rotating the parabolic arc about a vertical axis positioned on the concave side of the arc. It can be formed by a symmetrical arc or by the more widely used offset arc. This reflector is fed by multiple feed horns, each of which uses portions of the reflector surface used by adjacent horns. With multiple usage of the surface, a relatively compact structure can be obtained. The final surfaces with a parabolic arc are the elpar and hypar surfaces with an ellipse and a hyperboloid in the second principal plane. These reflectors are discussed in the next subsection. A recent spaceborne concept was suggested in Hoferer and Rahmat-Samii⁴² and Njoku et al.⁴³

15.4 OFFSET (SYMMETRIC) PARABOLOIDAL REFLECTORS

Since offset (symmetric) paraboloidal reflector antennas are still one of the most popular reflector antenna configurations, in this section the radiation characteristics of offset (symmetric) paraboloidal reflector antennas illuminated by a single-feed element (fixed-phase center) are presented with ample parametric studies. As is typical in any reflector design, there are too many almost-independent parameters that may be varied to achieve a particular design goal. For instance, to design an offset parabolic reflector, you must study the effects of such parameters as offset angles, illumination tapers, F/D ratios, locations and orientations of the feed, polarizations, etc., on such far-field pattern characteristics as scan loss, beamwidth, sidelobe level, cross-polarization level, efficiency, and so on. Obviously it is not possible to perform a comprehensive study of all these parameters; rather, attempts are made to present the most important reflector characteristics based on a few key parameters. The results of this section can therefore be used as a guideline for an initial design, which can then be refined by using computer programs and measured data.

Geometrical Parameters

The geometry of an offset parabolic reflector with focal length F , diameter D , and offset height H is shown in Figure 15-13. There are also other parameters used to characterize offset parabolic reflectors, which are defined next:

F = focal length

D = reflector diameter (diameter of the circular projected aperture)

H = offset height ($H = -D/2$ for symmetric reflectors)

d = offset height of the circular projected aperture center

$= D/2 + H$

ψ_U = angle subtended to the upper tip

$= 2 \tan^{-1}[(D + H)/2F]$

ψ_C = angle subtended to the center of the projected aperture

$= 2 \tan^{-1}[(D/2 + H)/2F]$

ψ_L = angle subtended to the lower tip

$= 2 \tan^{-1}[H/2F]$

ψ_B = angle subtended in the bisect direction

$= (\psi_U + \psi_L)/2$

ψ_S = half-angle subtended to the upper and lower tips

$= (\psi_U - \psi_L)/2$

ψ_{SC} = half-angle subtended to the right and left sides

$$= \sin^{-1} \left(\frac{D}{2} / \sqrt{\left(H + \frac{D}{2}\right)^2 + \left(\frac{D}{2}\right)^2 + \left[\frac{(D/2)^2 + (H + D/2)^2}{4F} - F\right]^2} \right)$$

D_p = parent parabola diameter

$= 2(D + H)$ for $H \geq -D/2$

In some cases F , ψ_B , and ψ_S are given, and from them D and H may be constructed. This can be done by using the following expressions:

$$D = 4F \sin \psi_S / (\cos \psi_B + \cos \psi_S)$$

$$H = 2F(\sin \psi_B - \sin \psi_S) / (\cos \psi_B + \cos \psi_S) \quad (15-16)$$

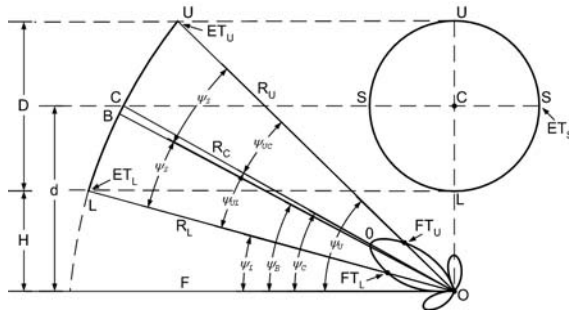


FIGURE 15-13 Geometrical parameters of an offset parabolic reflector ($H = -D/2$ for the symmetric case)

In designing offset reflectors any combination of the preceding parameters may be given, and the others may be constructed. In this subsection the parameters D , F/D , and H/D are used as independent parameters and the others are obtained. For example, Figures 15-14 and 15-15a through 15-15c show plots of F/D_p , ψ_{UC} , ψ_{CL} , ψ_{SC} , and $20 \log(R_C/R_{U,L,S})$ as functions of F/D for different H/D values. It is clear for symmetric reflectors that $F/D = F/D_p$ and $\psi_{UC} = \psi_{CL} = \psi_{SC}$, because $H/D = -0.5$. In the following subsections it will become apparent that these parameters play important roles to the extent that the far-field characteristics of reflectors are concerned. For example, different path losses at different edges (tips) of the reflector can cause different sidelobe levels in the far-field pattern cuts. In practice, H/D varies from 0.1 to 0.3 to provide clearance for the feed assemblies.

Idealized Feed Patterns

As mentioned earlier, a reflector's far-field pattern characteristics cannot be determined without a proper description of its feed patterns. A discussion of commonly used feed elements will be given in Chapter 13. However, to present parametric results concerning the performance of parabolic reflectors, idealized feed patterns, which have proven to be very

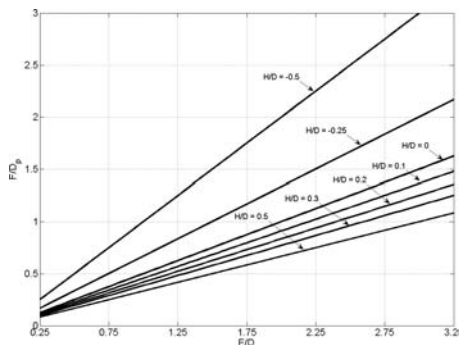


FIGURE 15-14 The ratio F/D_p for different values of H/D

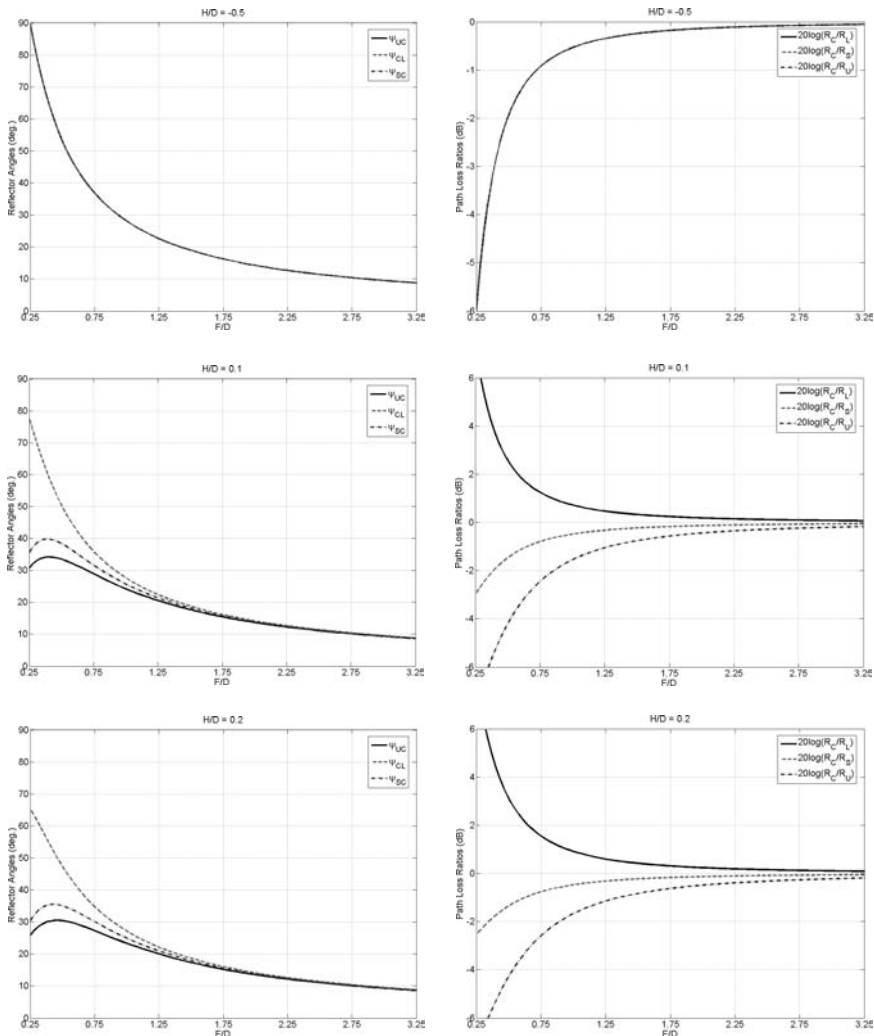


FIGURE 15-15 Reflector angles and path loss ratios, as defined in Figure 15-13, as a function of F/D : (a) for $H/D = -0.5$ (symmetric reflector case); (b) for $H/D = 0.1$ and (c) For $H/D = 0.2$

useful models, are used. For an idealized feed with a fixed phase center, its radiation pattern may be described as

$$\mathbf{E}(\mathbf{r}) = A_0 \begin{cases} \hat{\theta} C_E(\theta) \cos \phi - \hat{\phi} C_H(\theta) \sin \phi \\ \hat{\phi} C_E(\theta) \sin \phi + \hat{\theta} C_H(\theta) \cos \phi \end{cases} \frac{e^{-jkr}}{r}, \quad \begin{cases} \text{for } \hat{x} \text{ polarized} \\ \text{for } \hat{y} \text{ polarized} \end{cases} \quad (15-17)$$

where A_0 is a complex constant and

$$\begin{aligned} C_E(\theta) &= (\cos\theta)^{q_E} = E\text{-plane pattern} \\ C_H(\theta) &= (\cos\theta)^{q_H} = H\text{-plane pattern} \end{aligned} \quad (15-18)$$

for $0 \leq \theta \leq \pi/2$ and zero otherwise. There are other versions of this idealized feed pattern with a half angle representation that covers the entire angular range.

The shape of the pattern is controlled by q_E and q_H , which are determined by matching Eq. 15-18 to the given feed pattern. A proper superposition of the x and y polarized components in Eq. 15-17 also generates a circularly polarized field, namely,

$$\mathbf{E}(\mathbf{r}) = A_0 e^{j\tau\phi} [\hat{\theta} C_E(\theta) + \hat{\phi} j\tau C_H(\theta)] \frac{e^{-jkr}}{r} \quad (15-19)$$

where $\tau = +1$ for left-handed circular polarization, and $\tau = -1$ for right-handed circular polarization. Note that Eq. 15-19 represents a perfect circular polarized wave only in the main beam direction ($\theta = 0$). Away from this direction it is generally elliptically polarized unless $q_E = q_H$.

An attractive feature of Eq. 15-17 is that the directivity is derived in closed form:

$$D(\theta = 0) = \frac{2(2q_E + 1)(2q_H + 1)}{q_E + q_H + 1} \quad (15-20)$$

which consequently allows computation of the directivity of the reflector illuminated by this feed.⁴⁴ It must be noted that the back radiation of the feed pattern is ignored in deriving Eq. 15-20.

Edge and Feed Tapers

A widely employed parameter to characterize the effects of the feed element pattern on the far-field pattern of the reflector is the *edge taper*. In a broad sense this signifies the ratio of the field intensity at the reflector edge to the intensity at its center in decibels. Although this definition is unambiguous when it is applied to symmetric reflectors, it can become ambiguous for offset reflectors. For this reason another definition, referred to as *feed taper*, is defined.

Referring to Figure 15-13, the feed taper FT_U in the upper-tip direction is defined as

$$FT_U = 20 \log \left[\frac{C(\psi_U - \psi_C)}{C(0^\circ)} \right] \quad (15-21)$$

where 0° refers to the central direction (i.e., OC in Figure 15-13) and C denotes the feed pattern as defined in Eq. 15-18. Similar definitions for FT_L and FT_S can also be given at the lower and side angles by using ψ_C , ψ_L , and ψ_S , respectively. The edge taper ET_U at the upper tip may now be expressed as

$$ET_U = FT_U + 20 \log(R_C/R_U) \quad (15-22)$$

where R_C and R_U are the path lengths from the feed to the center and the upper tip of the reflector, respectively. The second term in Eq. 15-22 is also called the *path loss* term. Similar definitions can also be given for ET_L and ET_S . Note that for symmetric reflectors,

ET and FT only need to be defined at one edge (tip). For different values of F/D and F/D the path loss curves are shown in Figures 15-15a through 15-15c. It is worthwhile to add that for most cases of interest, ET_U , ET_L , and ET_S have nearly equal values. Since ET directly controls the reflector's aperture amplitude taper, it has a more dominant effect on the far-field pattern than FT does. It is not only the taper level that controls the reflector pattern but also the overall shape of the illumination distribution. In particular, the slopes of the illumination pattern at the reflector's edge can affect the sidelobe levels. The results of this section are primarily based on the $\cos^q(\theta)$ type illumination patterns.

Reflector Pattern Characteristics for On-Focus Feeds

Reflector pattern characteristics are discussed for beams generated by feeds located on the focus. Results are shown for beamwidths, sidelobe levels, first-null positions, cross-polarization levels, directivity efficiency, etc., as functions of the edge taper ET (or feed taper FT) and the reflector geometries. To simplify the presentation and limit the number of graphs, cases are considered in which the path losses are small (less than 0.5 dB) and, therefore, no substantial differences can be observed for path losses at the different tips of the reflector. For these cases, $\psi_{UC} \approx \psi_{CL} \approx \psi_{SC} \approx \psi_S$ and $FT \approx ET$. The reader should attempt to properly interpret the results when the path losses are substantial. This effect is clearly demonstrated in Figure 15-16, which displays far-field patterns for symmetric reflectors ($H/D = -0.5$) for the cases of $F/D = 0.4$ and $F/D = 2.0$, with an edge taper of $ET = -10$ dB. From Figure 15-15a it is concluded that the path losses are -2.86 dB and -0.13 dB for F/D s of 0.4 and 2.0, respectively. Nevertheless, the patterns are very similar, which indicates that $ET = -10$ dB is the controlling factor. Note that for these values of F/D the feed taper FT takes the values of -7.14 dB and -9.87 dB, respectively. In the latter case, $ET \approx FT$. Similar results are shown in Figure 15-17, where the feed taper FT is kept constant at -10 dB. The patterns for cases in which $F/D = 0.4$ and 2.0 show differences of up to 2.8 dB at the first sidelobe level, which is obviously a manifestation of the effects of the path loss differences.

For offset parabolic reflectors the far-field patterns in different cuts are, in general, different even when the feed has a symmetric pattern. Also, the pattern can be slightly asymmetric in the plane of offset depending on the F/D ratio (see Figure 15-18).⁴⁵ However, for the results shown here, F/D is large in order to reduce the path loss effects and asymmetry

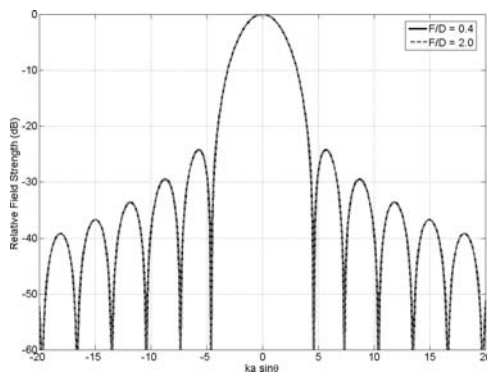


FIGURE 15-16 Far-field patterns of a symmetric reflector antenna with an edge taper illumination of $ET = -10$ dB and different path losses (different F/D values)

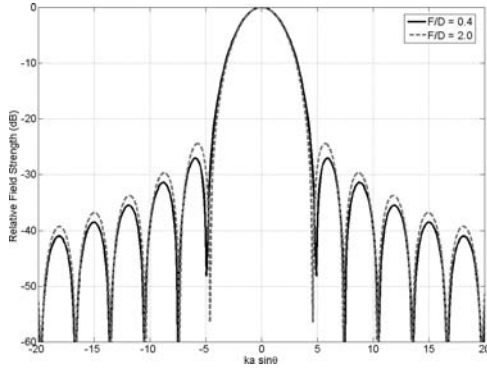


FIGURE 15-17 Far-field patterns of a symmetric reflector antenna with a feed taper illumination of $FT = -10$ dB and different path losses (different F/D values)

of the pattern. Figure 15-19 shows the half-beamwidth, first and second null positions, and first sidelobe positions as functions of the edge taper ET . It is worthwhile to mention that for edge tapers beyond -20 dB the pattern characteristics depend heavily on the actual feed pattern description, and results shown here are for $\cos^q(\theta)$ feed patterns. Also shown in Figure 15-20 are the first sidelobe levels and spillover and taper efficiencies as functions of edge tapers. From Figure 15-20 it is readily observed that the resulting efficiency $\eta = \eta_r \eta_t$ is maximized for edge tapers about -11 dB with a value of 81 percent. Some representative far-field patterns are shown in Figure 15-21 for different edge tapers. Again, it is obvious that for edge tapers in the neighborhood of $ET = -20$ dB, the first sidelobe starts to merge with the main beam, which results in a widened beam. For this level of edge taper the exact distribution of the feed pattern can have a significant effect on the pattern characteristics.

Another important reflector parameter is the level of generated cross-polarized field. This topic has been addressed in Love,³ Chu and Turrin,⁴⁶ and Ludwig⁴⁷ for both symmetric and

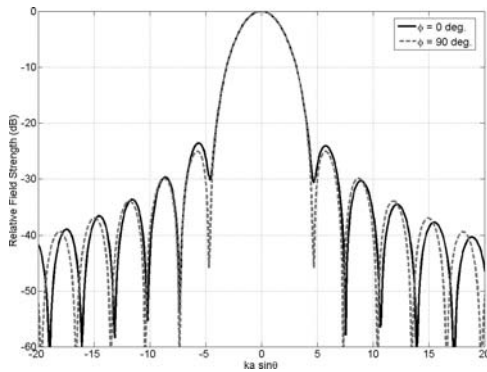


FIGURE 15-18 Patterns of an offset reflector with $H/D = 0.5$ ($F/D_p = 0.25$) illuminated by a symmetric feed with a feed taper of $FT = -10$ dB

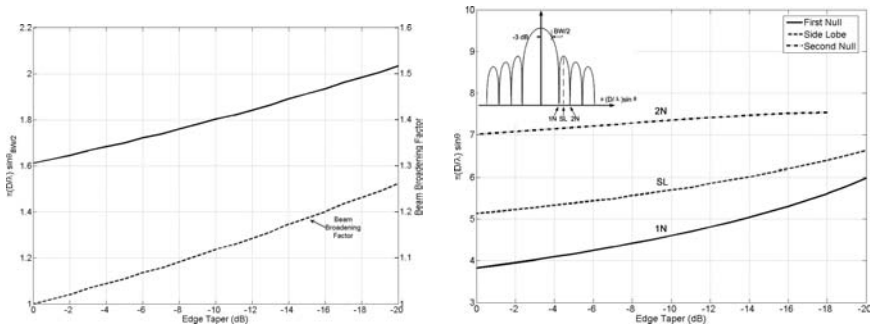


FIGURE 15-19 Half-power beamwidth, first-null, second-null, and sidelobe positions as functions of the edge taper, where D/λ is the diameter of the reflector in terms of the wavelength: (a) half-power beamwidth and (b) first and second nulls and sidelobe positions

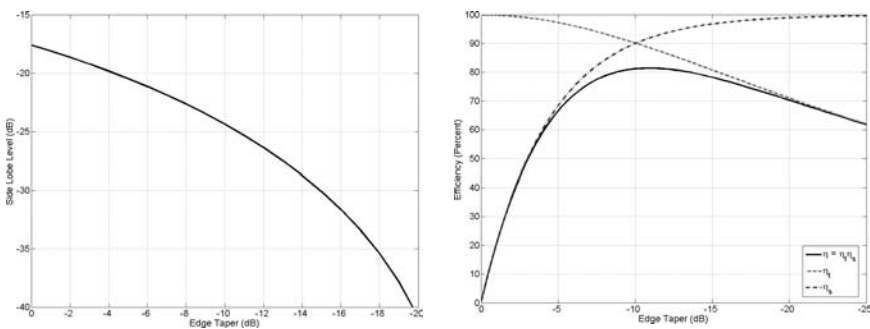


FIGURE 15-20 Sidelobe level, taper efficiency, spillover efficiency, and overall efficiency as functions of edge taper ET for $\cos^q(\theta)$ type feed patterns: (a) sidelobe level and (b) taper, spillover, and overall efficiencies

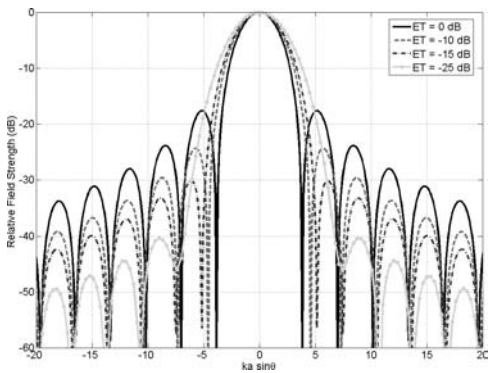


FIGURE 15-21 Reflector far-field patterns for different edge tapers

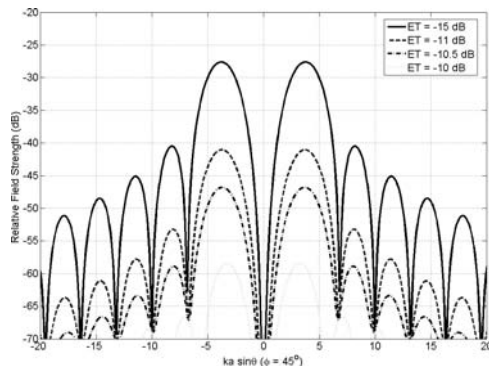


FIGURE 15-22 Cross-polarized field (normalized with respect to the main beam) of a symmetric reflector ($F/D = 0.3$) illuminated by a linearly polarized feed with unbalanced E - and H -patterns.

offset reflectors. Here, some representative cases are presented. There are many different cases that can be studied. Commonly the unbalanced feed refers to a feed with different E - and H -plane patterns. For the symmetric reflector the mechanism of generation of the cross-polarized field is related to the aperture field decomposition, and some typical patterns are given in Figure 15-22. Note that for a symmetric parabolic reflector illuminated by a linearly polarized feed, the maximum of cross-polarized fields occurs in the plane $\phi = 45^\circ$. As the feed becomes more balanced, the level of the cross-polarized field decreases substantially, as demonstrated by the results in Figure 15-22. It is worth mentioning that these results are dependent on the F/D ratio and the edge taper.

Next, the cross-polarization characteristics of offset parabolic reflectors are considered. In contrast to symmetric reflectors, which have very low levels of cross-polarized fields for balanced and linearly polarized feeds, offset parabolic reflectors can have a high level of cross polarization. Even for balanced feeds located at the focal point, levels of cross-polarized fields can be high, depending on the tilt angle of the feed axis with respect to the reflector axis for linearly polarized feeds. For example, Figure 15-23 shows the generation of cross-polarized fields for various feed-axis tilt angles for a fixed offset reflector configuration. In this example the feed pattern is chosen to be isotropic in order to show clearly the generation of the cross-polarized fields. It should be mentioned that for offset parabolic reflectors the cross-polarized field is predominantly observed in the plane $\phi = 90^\circ$ (normal to the plane of the offset). Clearly, in practice, to reduce spillover the feed axis is always tilted toward the center of the reflector; hence a high level of cross-polarized field should be expected. The levels of cross-polarized fields for different values of bisect angle ψ_B and the half-angle ψ_s (see Figure 15-23) can be generated.⁴⁶

When symmetric or offset parabolic reflectors are illuminated with balanced circularly polarized feeds, very low levels of cross-polarized fields result. For offset parabolic reflectors illuminated by circularly polarized feeds, however, an additional feature is observed, which is referred to as the *beam squint*. This means that the beam peak is shifted from the axis on the plane normal to the plane of offset. The amount of squint depends on the tilt angle of the feed axis and the reflector geometry. It can be shown that the following expression is a good approximation:

$$\sin \theta_s = \mp \frac{\sin \psi_B}{4\pi(F/\lambda)} \quad (15-23)$$

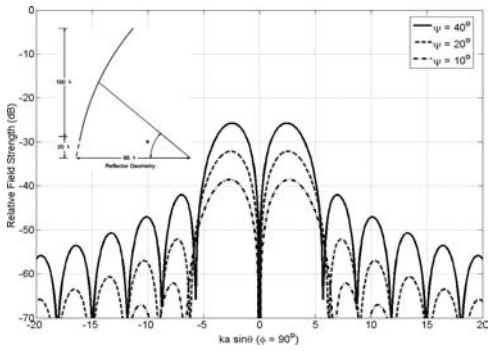


FIGURE 15-23 Cross-polarized fields in the plane $\varphi = 90^\circ$ (normal to the offset plane) for an offset parabolic reflector for various feed-axis tilt angles ψ and illuminated by a balanced linearly polarized feed

where \mp signs are for right and left circularly polarized cases and θ_s is the amount of squint. The amount of squint obtained from experimental data and diffraction analysis⁴⁶ agrees well with the approximate formula. Additional results are shown in Figure 15-24,⁴⁶ where again ψ_B is used to characterize the reflector.

Reflector Pattern Characteristics for Off-Focus Feeds

In many applications, such as the design of multiple and contour beam reflectors, it becomes necessary to illuminate the reflector with feeds positioned away from the reflector focal point. This feed displacement introduces phase aberration, which results in pattern distortion in terms of gain loss, sidelobe degradations, etc. In this subsection some of the key distortion characteristics of reflectors are presented for both the symmetric and offset parabolic reflectors and for both the axial and lateral feed defocusing.

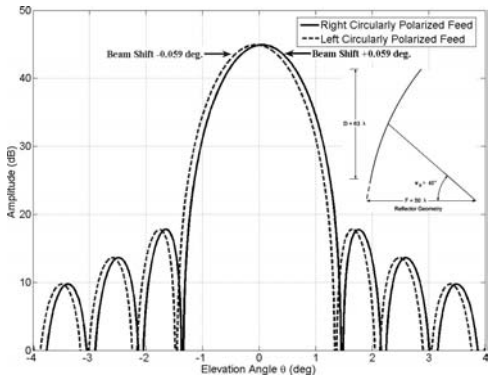


FIGURE 15-24 Squinted far-field patterns in the plane $\varphi = 90^\circ$ of an offset parabolic reflector illuminated by right and left circularly polarized feed located at its focal point

Axial Displacements For the symmetric reflector the results of feed axial defocusing for gain loss and patterns are shown in Figures 15-25 through 15-27. It is assumed that the feed provides a -10 -dB edge taper when it is located at the focal point. Results are demonstrated for feed displacements toward and away from the reflector. A small asymmetry can be observed. These results are dependent on the edge taper as discussed in Ingerson and Rusch.⁴⁸ Furthermore, for the cases where $\delta \ll F$ and the feed is an infinitesimal dipole with moment \mathbf{p} polarized in the $\hat{\mathbf{y}}$ direction, you can obtain the following approximate expression⁴⁸ for the field on axis:

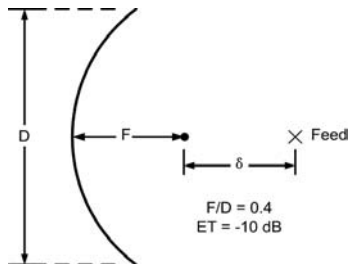


FIGURE 15-25 Relative on-axis field as a function of axial feed displacement

$$\mathbf{E}(R, 0, 0) = -\frac{jk^2 p}{4\pi\epsilon} (2kF) \hat{\mathbf{y}} \frac{\exp(-jkR)}{R} \frac{1}{(4F/D)^2 + 1} \exp\left[-jk\delta \frac{(4F/D)^2}{(4F/D)^2 + 1}\right] \frac{\sin \xi}{\xi} \quad (15-24)$$

$$\xi = \frac{2\pi(\delta/\lambda)}{(4F/D)^2 + 1} \quad (15-25)$$

The dominant effect of the axial defocusing is the generation of the quadratic phase error across the reflector aperture. The axial field (see Eq. 15-24) becomes virtually zero for values of $\xi = \pm\pi, \pm2\pi, \pm3\pi$, etc., which results in

$$\frac{\delta}{\lambda} = \pm \frac{n}{2} [(4F/D)^2 + 1], \quad n = 1, 2, 3, \dots \quad (15-26)$$

For these values the beam widens considerably and may also be bifurcated (see Figures 15-26 and 15-27). You may have noticed a resemblance between these patterns and those resultant from the field of an aperture in the Fresnel zone.

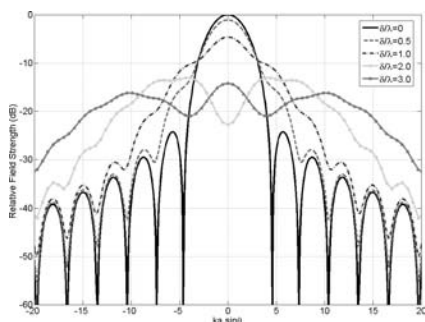


FIGURE 15-26 Reflector far-field patterns for $\varphi = 0^\circ$ as a function of axial feed displacement away from the reflector

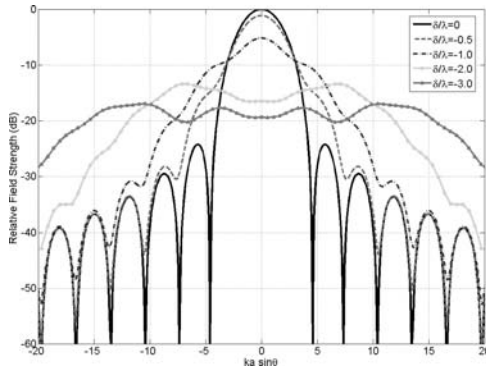


FIGURE 15-27 Reflector far-field patterns for $\varphi = 0^\circ$ as a function of axial feed displacement toward the reflector

Lateral Displacements For a parabolic reflector, lateral feed displacements result in scanned beams. It is well known that these reflectors have limited scan capability that strongly depends on F/D and F/D_p ratios for symmetric reflectors. An important parameter in dealing with scanned beams is the beam deviation factor (BDF), which is defined as

$$BDF = \frac{\theta_B}{\theta_F} \quad (15-27)$$

where θ_B and θ_F are the beam scan angle and feed tilt angle, respectively, as shown in Figure 15-28. The BDF has a strong dependence on F/D and weakly depends on the edge taper and δ/F (δ is the lateral feed displacement). A good approximation for BDF is⁴⁹

$$BDF = \frac{\sin^{-1}(\delta/F) \{ [1 + k(D/4F)^2] / [1 + (D/4F)^2] \}}{\tan^{-1}(\delta/F)} \quad (15-28)$$

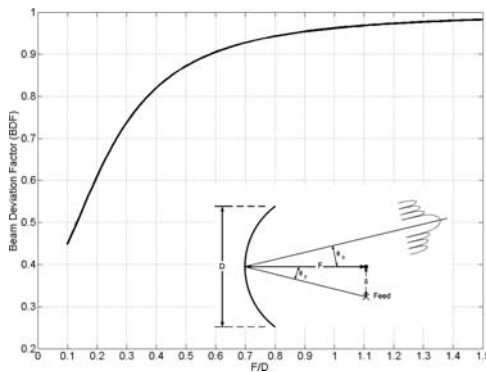


FIGURE 15-28 Beam deviation factor ($BDF = \theta_B/\theta_F$) as a function of the F/D ratio for symmetric reflectors ($ET = -10$ dB)

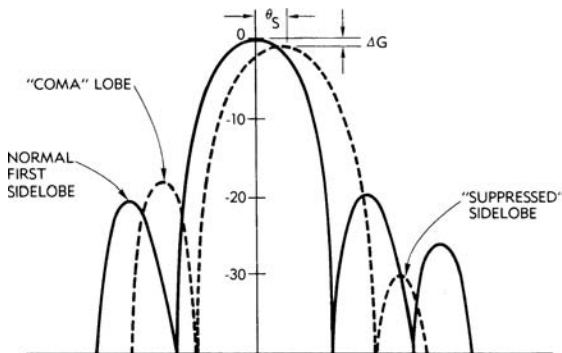


FIGURE 15-29 Radiation patterns for focused and laterally defocused feeds

where $0 < k < 1$ and $k = 0.36$ provide very accurate results when the BDF is compared with experimental and diffraction analysis data. For small δ/F , Eq. 15-28 may be simplified to give

$$BDF_0 = \frac{1 + k(D/4F)^2}{1 + (D/4F)^2} \quad (15-29)$$

which is independent of δ/F . Figure 15-28 shows the BDF as a function of F/D and for $k = 0.36$. For large feed displacements you may have to resort to diffraction analysis results in order to obtain more accurate values for the BDF.

Due to the phase aberration introduced by defocused feeds, the far-field pattern can be substantially degraded, as shown in Figure 15-29. This degradation depends very strongly on the F/D ratio and the angle of scan in terms of number of beamwidths scanned (bmws). Some approximate formulas are available for predicting the peak gain loss as a function of beamwidths scanned.⁵⁰ In many applications, however, the knowledge of peak gain loss is not sufficient, and you must know the overall degradation effects on the far-field pattern. For this reason many selective but representative cases are given here to provide a clear picture of the pattern degradation. First, an $F/D = 0.4$ symmetric parabolic reflector illuminated by a $\cos^q(\theta)$ type feed with $ET = -10$ dB is considered. Far-field patterns for different numbers of beamwidths scanned are shown in Figure 15-30. It is important to note that the patterns are plotted versus the universal parameter $ka \sin \theta_p = \pi(D/\lambda) \sin \theta_p$, in which θ_p is the angle measured from the axis passing through the peak of the beam that is directed in the $\theta_b = BDF \tan^{-1}(\delta/F)$ direction. The peak values of these patterns are normalized with respect to the peak value of the nonscanned beam. Note that for large scan angles, $BDF = 0.82$ (for $F/D = 0.4$) does not predict the location of the beam maximum accurately. Furthermore, the feed axis is kept parallel to the reflector axis and has not been tilted toward the reflector center. This condition is more practical when feeds are used in planar array in contour and multiple beam applications. For directive feeds with large displacements, however, it may be necessary to tilt the feed toward the reflector center in order to reduce the amount of spillover. In this case the feeds may be arrayed on the spherical surface rather than the planar surface. Figure 15-30 clearly demonstrates how rapidly the reflector pattern can degrade for scan angles beyond two beamwidths scanned when $F/D = 0.4$.

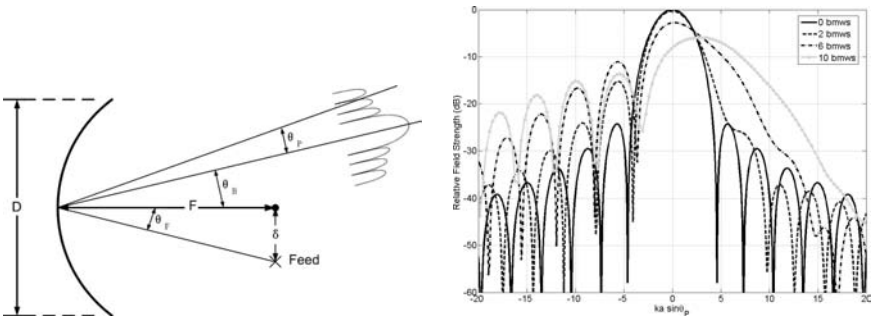


FIGURE 15-30 Symmetric reflector far-field patterns for $\phi = 0^\circ$ plane as a function of lateral feed displacement in terms of the number of beamwidths scanned ($F/D = 0.4$, $ET = -10$ dB, $BDF = 0.82$): (a) reflector parameters and (b) far-field patterns

One simple way to improve this very limited scan capability is to increase the F/D ratio. Figure 15-31 shows the far-field patterns for $F/D = 1.0$ as a function of beamwidths scanned. Note that for this value of F/D , beams with much larger scan angles can be generated with adequate characteristics. The prime drawbacks of employing large F/D values are the larger structural size and the need for more directive feeds to provide the required edge taper. Finally, Figure 15-32 shows the gain loss curves for different values of F/D as a function of the beamwidths scanned (bmws). Although most of the results are shown for the case of $ET = -10$ dB, similar observations can be made for other edge tapers. For example, Figure 15-33 shows the results of experimental data.⁵¹ Excellent agreement has been observed as far as the effects of the feed displacements are concerned. There are, however, differences in the third sidelobe levels due to the fact that the experimental feed patterns are not exactly modeled by the $\cos^q(\theta)$ type pattern.

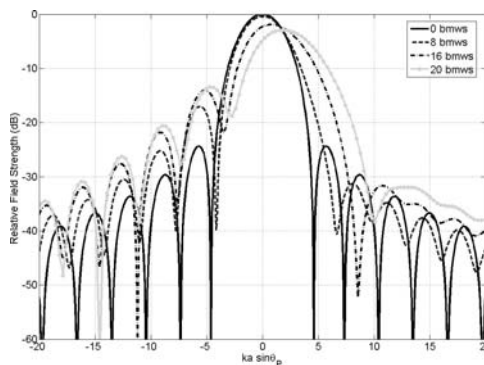


FIGURE 15-31 Symmetric reflector far-field patterns for $\phi = 0^\circ$ plane as a function of lateral feed displacement in terms of the number of beamwidths scanned ($F/D = 1.0$, $ET = -10$ dB, $BDF = 0.966$)

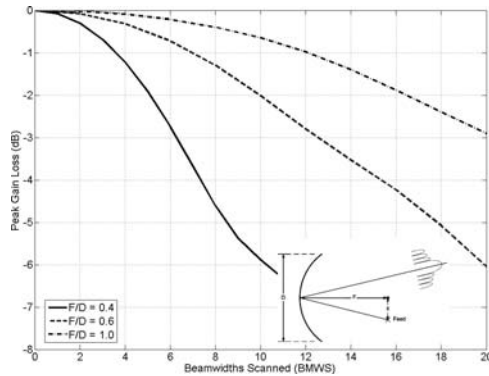


FIGURE 15-32 Peak gain loss of symmetric parabolic reflectors as a function of beamwidths scanned for different F/D ratios

The previously mentioned results for symmetric reflectors only demonstrate the scan characteristics while ignoring the blockage effects, which could become very severe if large feed arrays are used to illuminate reflectors for multiple and contour beam applications. To overcome these blockage effects, the designers of satellite communications systems frequently employ offset reflectors that have scan characteristics as discussed here. For offset reflectors the beam deviation factor is computed as shown in Figure 15-34. Many studies have shown that the F/D_p ratio characterizes the offset reflector patterns better than the F/D ratio. Far-field patterns as a function of the number of beamwidths scanned are shown in Figure 15-35 for $F/D_p = 0.4$ and $ET = -10$ dB. For this case $F/D = 0.96$ ($F = 96\lambda$, $D = 100\lambda$, $H/D = 0.2$), which is almost two and a half times larger than F/D_p . Notice the similarity between these patterns and those of the symmetric reflector with $F/D = 0.4$ (Figure 15-30b). Similar results are also shown for an offset reflector with $F/D_p = 1.0$ and $F/D = 2.4$. Figure 15-36 shows the patterns for the cases in which the feed

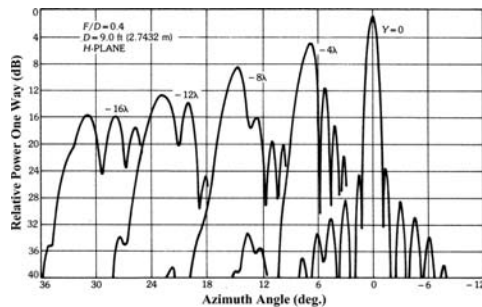


FIGURE 15-33 Measured secondary patterns as a function of lateral primary feed displacement for a symmetric parabolic reflector⁵¹

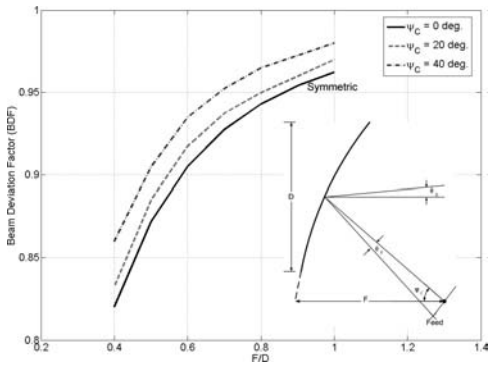


FIGURE 15-34 Beam deviation factor versus F/D for different offset angles ψ_C

direction is kept fixed as the feed is displaced in the plane orthogonal to the line joining the focal point with the reflector center. For this large value of F/D_p , the feed is very directive in order to provide the -10 -dB edge taper and, therefore, for large scans the reflector would be very poorly illuminated. Results for the cases in which feed is tilted toward the reflector center while it is displaced in the focal plane are shown in Figure 15-37. Notice that a marked improvement can be observed in the patterns. Finally, the peak gain losses as a function of beamwidths scanned and for different values of F/D_p are plotted in Figure 15-38. It is worthwhile to mention that there are a considerable number of ongoing attempts to improve the scan performance of single-reflector antennas by employing the concept of conjugate matched focal-plane feed arrays. This approach may also be utilized to overcome the deterministic reflector surface distortions.

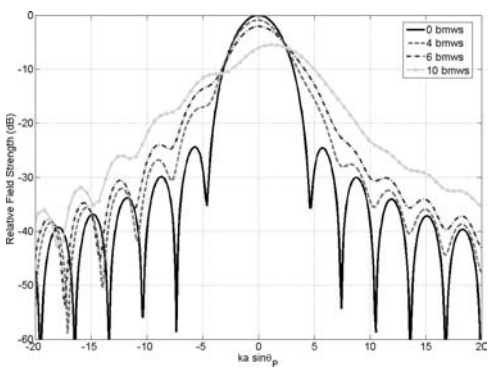


FIGURE 15-35 Offset reflector far-field patterns as a function of lateral feed displacement in terms of the number of beamwidths scanned ($F/D_p = 0.4$, $F/D = 0.96$, $ET = -10$ dB, $BDF = 0.983$, $\varphi = 0^\circ$)

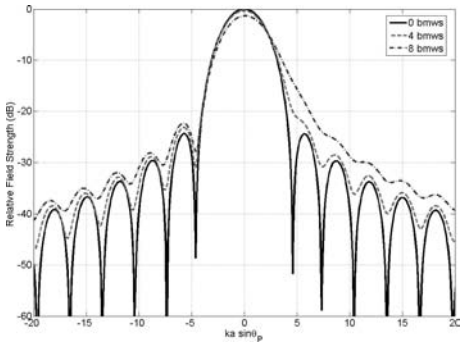


FIGURE 15-36 Offset reflector far-field patterns as a function of lateral feed displacement in terms of the number of beamwidths scanned ($F/D_p = 1.0$, $F/D = 2.4$, $ET = -10$ dB, $BDF = 0.998$, $\varphi = 0^\circ$)

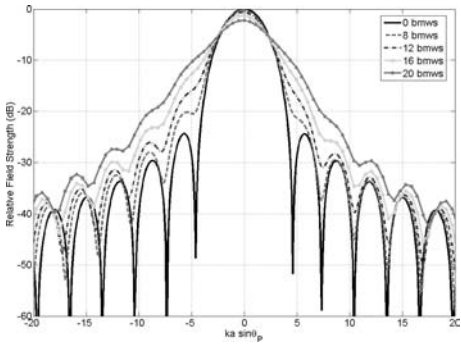


FIGURE 15-37 Offset reflector far-field patterns for tilted feed as a function of lateral feed displacement in terms of the number of beamwidths scanned ($F/D_p = 1.0$, $F/D = 2.4$, $ET = -10$ dB, $BDF = 0.998$, $\varphi = 0^\circ$)

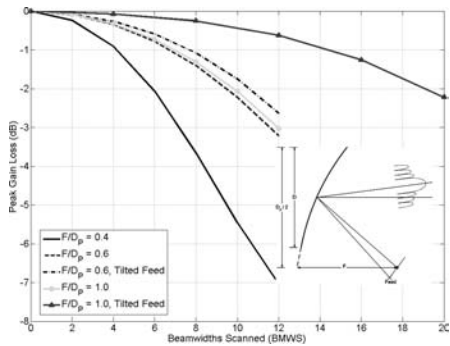


FIGURE 15-38 Peak gain loss of offset parabolic reflectors as a function of the number of beamwidths scanned for different F/D_p ratios

15.5 REFLECTORS OTHER THAN PLANE OR PARABOLIC

Some key feature of popular and non-parabolic reflector antennas are discussed in this section. Attempts have been made to provide design curves and an overview of their radiation characteristics.

General Surface

There are a number of useful surfaces in addition to the plane and the surfaces with a parabolic arc. The coordinates of these surfaces are determined by the required phase and amplitude at points on the reflected wavefront. The phase is associated with the length of the incident and reflected ray paths,⁵² and the amplitudes depend upon the density of the rays.⁵³ The most widespread use of general surfaces has been in shaped beams and Cassegrain antennas, but other useful surfaces are included here.

Sphere

The first surface considered is the sphere (see Figure 15-39), which was shown in Kelleher⁵² to have the phase error $2(1 - a^2)^{1/2} + 2(da^2 - 1)$, where d is the distance from the sphere center to the feed and a is the distance from the centerline to the spherical surface. The sphere has unit radius. A practical sphere is limited in extent. If an attempt is made to scan it to one side, the phase errors do not remain the same, but increase as the beam is moved off axis. That is, the value of r in the expression for phase error must increase on one side and decrease on the other. Consider a plane through the sphere containing the axis. The phase-error expression can be plotted in Figure 15-40 for a number of values of the feed position d . The reflector edges move across the phase-error curves as the feed horn is

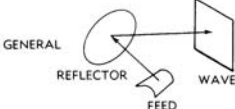

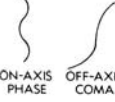
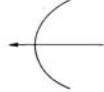

REFLECTOR	USES	REMARKS
<div>GENERAL </div>	REFLECTOR ANALYSIS & SYNTHESIS	
<div>SPHERICAL </div>	SCANNING & MULTIPLE BEAMS	<div></div>
<div>HYPERBOLOID OR ELLIPSOID </div>	CASSEGRAIN SUBREFLECTOR	SEE HYPAR/ELPAR FOR FAN BEAM (FIG 17-8)
<div>CONICAL </div>	LOW-COST REFLECTOR (NARROW BEAM)	PARABOLIC-ARC SUBREFLECTOR FOR CONICAL WAVEFRONT

FIGURE 15-39 Reflectors other than parabolic

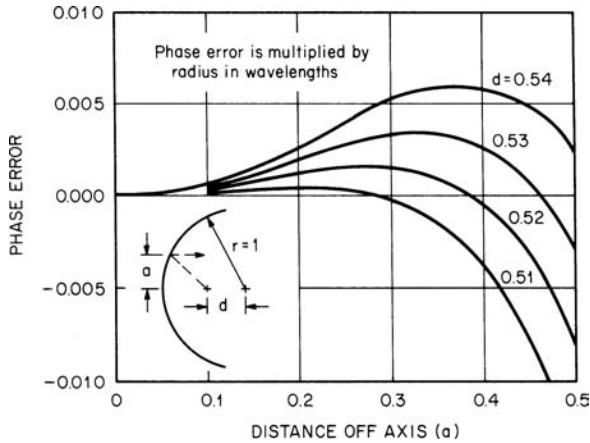


FIGURE 15-40 Phase error in a spherical reflector

moved off axis. The error will decrease slightly on one side and increase very rapidly on the other side. A careful analysis of the error will show that it has a strong cubic term (coma), just as the paraboloid has a strong cubic term. The sphere and the paraboloid have similar off-axis characteristics.

If the spherical reflector can have an efficiency on the order of 10 percent, it will give patterns that do not change with the beam moved off axis.⁵⁴ In this instance, the sphere is fed from a large feed horn that confines its radiation to a small portion of the surface. The reflector edges that were the problem in the normal sphere are then eliminated. In effect, this solution uses a very-long-length system in which the paraboloid and the sphere become very much alike.

The design of the sphere is based upon the phase-error formula. The error curve is used out to the point at which the phase error is zero. An argument can be made that some phase error beyond this point can be accepted. A tapered illumination from the feed will minimize these errors. In practical design, however, the phase error increases so fast that only a marginal improvement is achieved by considering the feed illumination. Some fine-tuning can be done with illumination, but the basic design depends upon the simple use of the two points on the phase-error curves, namely, the maximum value and the edge (zero) value that defines the maximum value of the aperture radius r . A simple ratio, namely, maximum phase error divided by r , is used. Let the phase error be $\lambda/16$ and the radius be 10λ (aperture diameter of 20λ). The ratio is 160:1. The design consists of selecting a specific curve using the parameter d that gives a ratio of 160:1. Note that the y coordinate in the figure is greatly expanded compared with the r coordinate.

Point-Source to Point-Source Reflectors

For some applications, it is desirable to produce a virtual source from a given real source. For example, it may be desirable to position a real source at the vertex of a paraboloidal reflector and utilize some surface between the vertex and the focal point that images the real source into a virtual source located at the focal point. The resulting antenna system, a counterpart of the optical Cassegrain system, will then have a focused beam, since the paraboloid is energized by a source that is apparently at the focal point.

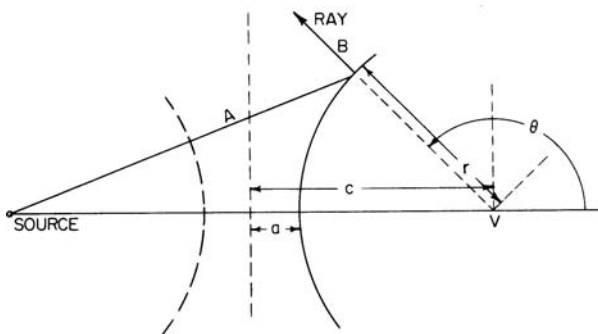


FIGURE 15-41 Geometry of hyperbola reflection

It is first obvious that the reflector surface desired must be symmetrical about the axis joining the two sources. If a cross section of the reflector is known, the complete surface can be formed by rotating this reflector arc about the axis. Figure 15-41 indicates the geometry of the two-dimensional problem. The reflected rays should appear to originate at the point v , or, in other words, rays from the real source shall be reflected to form a spherical wavefront with its center at v . Since in the figure we require that $A + B$ be a constant and $r + B$ should be a constant, we find that the surface is determined by the fact that $A - r$ is a constant. From analytic geometry, it is known that the locus of points satisfying this condition is a hyperbola. The curve can then be written as

$$r = \frac{a^2 - c^2}{a \pm c \cos \theta} \quad (15-30)$$

The quantities here are indicated in the figure. The plus-or-minus sign corresponds to two possible reflector surfaces, one concave and one convex.

Another possible reflector for translating a point source into a point source is an ellipsoid. The ellipse cross section is given by

$$r = \frac{a^2 - c^2}{a \pm c \cos \theta} \quad (15-31)$$

where the quantities are the same as in Figure 15-41.

For application to the paraboloidal reflector mentioned in the first paragraph of this subsection, three surfaces, namely, the ellipsoid and the two hyperboloids, are possible. For practical application, the ellipsoid, which must be mounted at a greater distance from the reflector, is not satisfactory. Of the two hyperboloid surfaces, the one that can be mounted farther from the reflector illuminates typical reflectors better than the nearer, which requires a paraboloid with large F/D ratio. When the real source is moved from one focal point of the hyperboloid system, the image source will move away from the other focal point but will, at the same time, become distorted. Any great motion of the real source produces a poor image.

Point-Source to Line-Source Reflectors

It has been mentioned that the parabolic-cylinder reflection will convert a point source into a line source. However, this property is not unique for this reflector. Other surfaces⁵⁵

capable of accomplishing the same result have been designated as hyper and elpar surfaces. These have principal plane sections that are hyperbolas, parabolas, and ellipses. It can be seen that in the plane of the parabola, rays from the feed source can be focused to be parallel, whereas in the plane of a hyperbola (or an ellipse), the rays can be reflected so that they appear to come from a virtual-source point. This does not prove that the reflected wave is a cylindrical wavefront, but indicates only that the cross sections of this wave in the two principal planes correspond to those of a cylindrical wave. A general reflector can be obtained by assuming that the point source lies at the coordinate value $(d, 0, 0)$ and that the line source coincides with the z axis as in Figure 15-42. By taking a general ray not lying in either principal plane and requiring that it have a constant path length between the point source and the line source, it can be shown that the reflector surface is given by the following expression:

$$\sqrt{(r \sin \theta)^2 + (r \cos \theta - d)^2 + z^2} + r = 2x_0 - d \quad (15-32)$$

or

$$\sqrt{(x-d)^2 + y^2 + z^2} + \sqrt{x^2 + y^2} = 2x_0 - d \quad (15-33)$$

All quantities in this expression are shown in Figure 15-42. If you set θ equal to a constant, it can be shown that all such sections of the reflector are parabolas, whereas setting z equal to a constant yields ellipses in those planes.

A reflector directly related to that indicated earlier involves cross sections that are hyperbolas rather than ellipses. The mathematics is similar. Using the quantities of Figure 15-42, we have the expression

$$\sqrt{(r \sin \theta)^2 + (r \cos \theta - d)^2 + z^2} - r = \pm(2x_0 - d) \quad (15-34)$$

The plus-or-minus signs indicate two possible surfaces, one concave and one convex. Here it can be shown that the phase θ constant intersect at the surface in parabolas and the planes z constant intersect it in hyperbolas.

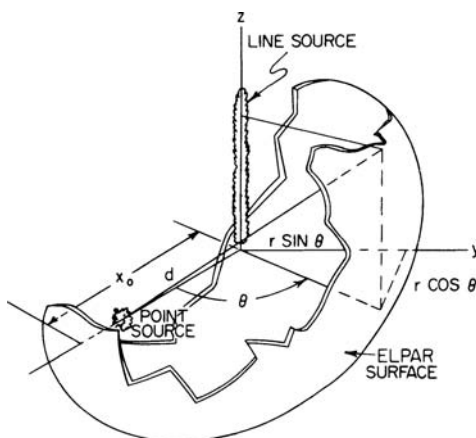


FIGURE 15-42 Geometry of elpar surface

Conical Reflector⁵⁶

The conical reflector is excellent in applications in which the complexity of a doubly curved large aperture is to be avoided. A simple feed with spherical wavefront is reflected from a unique surface into a conical wave that is reflected from the conical surface as a plane wave. Figure 15-43 shows a cross section of the three elements that are the basis of this system. The elements are the small horn, the small parabolic-arc reflector, and the conical reflector, which in the figure becomes a straight line. The complete system is obtained by rotating the cross sections about the axis. It can be seen that the tilted straight line when rotated about the axis becomes a cone with the center area open. The parabolic arc when rotated produces an unusual surface with a pointed vertex like a cone. The feed structure (small aperture and parabolic arc) blocks the center of the cone to give large sidelobes. An offset version of this system will eliminate the blocking, but the additional mechanical complexity reduces the advantage of using the simple cone.

The design of the conical reflector is based upon two angles: (1) the cone angle and (2) the angle that defines the subreflector. The cone angle is found to be 26° ; the angle defining the subreflector is about 50° . The equation of the generating curve of the subreflector is $\rho = 2F/[1 + \cos(52^\circ - \theta)]$. The zero coordinate of this curve coincides with the phase center of the feed. Ludwig⁵⁶ gives detailed design data that show that the optimum configuration has a usable annular ring of aperture with width 1.25 times the radius of the parabolic-arc reflector. Experimental measurements showed a gain in excess of 40 dBi with sidelobes of -12 dB.

The conical reflector can also be used as a magnifier system in which a smaller aperture radiates a plane wave toward a conical subreflector of equal aperture. The reflected conical wavefront then is used to feed the same conical reflector of Figure 15-43. The concept is essentially the same as that of the figure except that the initial wavefront is a plane rather than the spherical wave from the horn antenna. The magnification gives an increased directivity of the order of 2.5:1.

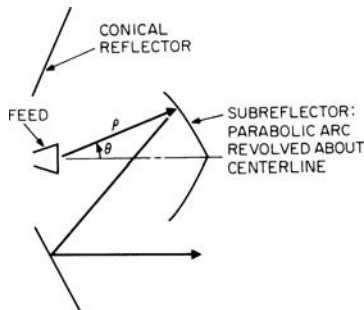


FIGURE 15-43 Conical reflector

Shaped-Beam, Singly Curved Reflectors

This problem is essentially two-dimensional, so that ray paths can be restricted to a plane and the cylindrical-reflector surface can be obtained from its plane-curve generator. The basic concepts for the design of a singly curved reflector are simple; the desired shaped-beam pattern is obtained point by point from the primary-feed pattern. The problem presented is that of forming the reflector to make any particular segment of primary-pattern energy appear at the desired point in space. Two basic conditions must be satisfied. The first is that of the energy correspondence between primary $I(\phi)$ and secondary pattern $P(\theta)$, which can be expressed as

$$\frac{\int_{\theta_1}^{\theta} P(\theta) d\theta}{\int_{\theta_1}^{\theta_2} P(\theta) d\theta} = \frac{\int_{\phi_1}^{\phi} I(\phi) d\phi}{\int_{\phi_1}^{\phi_2} I(\phi) d\phi} \quad (15-35)$$

where ϕ and θ are the primary- and secondary-pattern angles and the subscript values correspond to the reflector limits. The second condition relates the angle of incidence to the radius vector defining the surface:

$$\frac{\rho'}{\rho} = \tan i = \tan \frac{\phi - \theta}{2} \quad (15-36)$$

The relationship between θ and ϕ obtained from the first expression can be substituted in the second expression so that a differential equation is obtained between the radius vector ρ and the feed angle ϕ ; this relationship defines the desired reflector curve.

The most difficult part of the shaped-beam problem is that of computing the desired curve from the preceding expressions. Quite often the first expression must be solved graphically so that the differential equation also requires a graphic solution. If the functions $P(\theta)$ and $I(\phi)$ are integrable, then the first expression can be solved for $\theta(\phi)$ and the second expression becomes

$$\ln \frac{\rho}{\rho_1} = \int_{\phi_1}^{\phi} \tan \frac{\phi - \theta(\phi)}{2} d\phi \quad (15-37)$$

and the integral is evaluated by numerical methods. If the pattern functions are not integrable directly, numerical methods are used. First, a plot is made of the pattern functions. Next, the four integrals in the first expression are evaluated by obtaining the corresponding areas under the curves. For the integrals with variable upper limits, many different values of the area must be found, corresponding to different values of the variable. With this information, it is possible to obtain the two curves of Figure 15-44. These curves are used to obtain the desired value of θ for any given value of ϕ . With this known, $\tan \{[\phi - \theta(\phi)]/2\}$ is plotted and the integral evaluated so that $\rho(\phi)$, the desired curve, can be obtained.

Doubly Curved Reflectors⁵⁷

The problem of beam shaping becomes somewhat more complicated when the reflector is to shape the beam in one principal plane and focus it in the other plane. Since the problem is a 3D one, the possibility also exists that the feed might not produce a simple spherical wavefront but might also yield a cylindrical surface, a football-type surface, or any of a number of other surfaces. Fortunately, all feed systems of interest have a circular wavefront in the plane of beam shaping, so that techniques similar to that employed in the preceding subsection can be utilized to determine the surface cross section in that plane. The existing techniques consist of forming the reflector surface from the plane-curve cross section, which serves as a spine, and a series of other plane curves that are attached as ribs to the spine. Some question has arisen regarding this general technique, but it has been found to produce satisfactory experimental results for cases in which beam shaping is desired over a limited angle. This problem may be studied more carefully for application to beam shaping over wider angles.

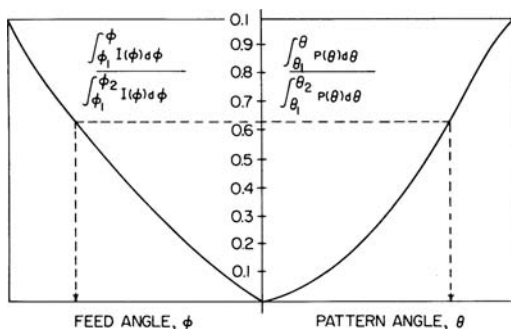


FIGURE 15-44 Relationship between feed angle and pattern angle in a shaped reflector

The plane curve, just identified as a spine, has been called the central-section curve. This curve is found in a manner quite similar to that described in the preceding subsection. The design equations for the feed with spherical wavefront are obtained using the same quantities as before:

$$\frac{\int_{\theta_1}^{\theta} P(\theta) d\theta}{\int_{\theta_1}^{\theta_2} P(\theta) d\theta} = \frac{\int_{\phi_1}^{\phi} \frac{I(\phi)}{\rho} d\phi}{\int_{\phi_1}^{\phi_2} \frac{I(\phi)}{\rho} d\phi} \quad \frac{\rho'}{\rho} = \tan \frac{\phi - \theta}{2} \quad (15-38)$$

The major difference between these expressions and the previous ones occurs in the presence of ρ under the integral sign. The integration can be carried out as before if first ρ is assumed to be a constant over the shaped portion of the central-section curve; then follow a parabolic arc [$\rho = \sec^2(\phi/2)$] over the region that produces the main beam. With this assumption, the procedure is identical with that of the last subsection for obtaining the function $\rho(\phi)$. This function represents a closer approximation to the correct value than the original assumption. When it is used as the value under the integral sign and the entire procedure is repeated, a new value of $\rho(\phi)$ that is as close to the true value as necessary in any practical case is obtained.

A discussion of the rib curves will indicate the limitations in the present techniques. To obtain the ribs, it is customary to consider an incoming bundle of rays from the secondary-pattern angle θ . For beam shaping in the vertical plane, this bundle of rays is confined to a plane making an angle θ with the horizontal plane. It is required that the reflector surface direct these rays to the feed position. Figure 15-45 shows two parallel rays in this plane that intersect the reflector in the curve AA' . The rays between the points A and A' are reflected to lie on a conical surface with apex at the feed.

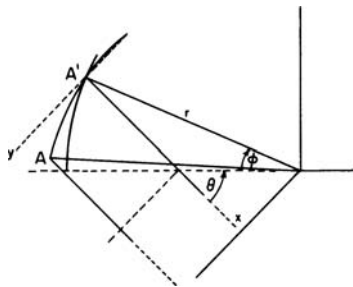


FIGURE 15-45 Geometry of a 3D shaped reflector

If we considered the transmit case and a plane sheet of rays emanating from the feedpoint, they would intersect the reflector in a curve AA' , and the reflected rays would lie on a cylindrical surface tilted at an angle of θ with the horizontal plane. It should be apparent that although they have the same endpoints, the plane curves AA' are not the same for the transmit and receive cases. This can be understood since it is known that if focusing is desired, the reflector surface in the neighborhood of the points of ray incidence must be a paraboloid. The plane curve cut by this paraboloid will be different if we consider the intersection caused by the transmit plane passing through a focal point and that caused by the receive plane inclined at an angle θ to the horizontal. It is difficult to make a choice between the two curves, although the great majority of existing efforts have chosen the receiving-plane case. A compromise solution might be the plane bisecting the angle formed by the transmitting and receiving planes. The parabolic curve for the receive case mentioned earlier is

$$y^2 = 4z\rho \cos^2 \frac{\theta + \phi}{2} \quad (15-39)$$

Here z is the coordinate measured from the point A' away from the reflector, and y is the orthogonal coordinate. It should be noted that each of the rib curves is a parabola lying in a different plane; when the reflector surface is formed, the planes of various parabola ribs must be inclined at the corresponding angles.

All the previous expressions yield only normalized coordinates for the surface; these must be converted to usable dimensions before the reflector can be constructed. It is desirable to make the central-section curve as large as possible in order to minimize the diffraction effects on the shaped-beam pattern. In the other plane, the reflector size is chosen to produce the desired beamwidth. The feed angles associated with the reflector are chosen in a manner similar to those of the simple paraboloid. It is desirable that the aperture illumination be 10 dB down at the reflector edges.

The most advanced technique⁵⁸ for defining the reflector for beam shaping permits beam control in both principal planes. A two-variable generalized far-field pattern is used to define the reflector surface. The generalization is achieved by replacing the ray-to-ray correspondence by a curve of incident ray directions related to a curve of reflected ray directions.

The caution in the use of any ray relationship is that it ignores diffraction effects. Therefore the reflector defined by geometrical optics should be checked by defining the phase and amplitude across the reflector aperture and integrating for the far-field patterns.

Shielded-Aperture Reflectors

In some applications it is desirable to have very low sidelobes from a pencil-beam reflector. In this instance, considerable improvement can be obtained by the use of metal shielding around the reflector aperture (see Figure 15-46). The typical sidelobes are at about 0 dBi, which for most reflectors represents a value of the order of -30 to -40 dB below the peak gain. With a shielding technique, the far-out sidelobes can be reduced to -80 dB.

This simplest approach to shielding the reflector is a cylindrical "shroud," or tunnel, of metal around the edge of a circular reflector (see Figure 15-46). If the aperture is elliptical in cross section, an elliptical cylinder can be used. The tunnel forms a waveguide of large cross section with the paraboloid at one end. The radiated energy has low levels of current on the outside edge of the tunnel and so has little radiation at wide angles from the beam. There is energy in the waveguide that leaves the feed horn to reflect from the inside of the cylinder. This energy is then reflected by the paraboloid into angles away from the axis.

An improved shielding system that eliminates the reflection of the feed-horn energy is shown in Figure 15-47 as the *horn reflector*. The two versions employ a pyramidal horn and


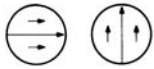






SHIELDED APERTURES	APERTURE FIELDS	REMARKS
CYLINDRICAL SHROUD 		REFLECTIONS FROM CYLINDRICAL SHROUD
HORN REFLECTOR 		HIGH E-PLANE ILLUMINATION
CONICAL HORN REFLECTOR 		COSTLY FABRICATION
FEED HORN CASSHORN 		CONTROLABLE ILLUMINATION FAN BEAM POSSIBLE

FIGURE 15-46 Shielded-aperture reflectors

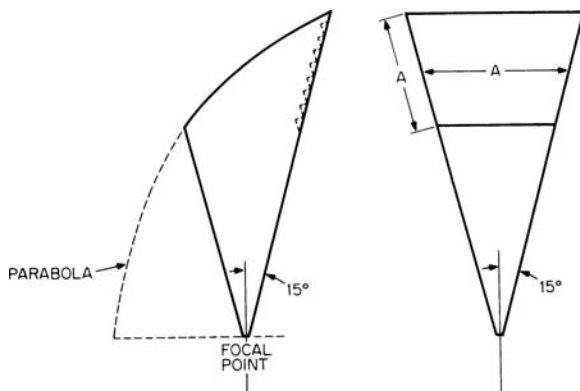


FIGURE 15-47 Geometry of horn reflector

a circular-waveguide horn. These antennas are discussed in other chapters. The variation among cornucopia horns is associated with the variation in aperture illumination. In many applications, the low sidelobe capability is associated primarily with the horizontal plane where interfering sources are found. The circular aperture will give reduced lobes, and the diagonal aperture (see discussion of aperture illumination in the preceding subsection) will give the best performance. The references for experimental cornucopia design are given in other chapters.

The final shielded horn to be considered is the casshorn,⁵⁹ which uses a secondary reflector to replace the horn structure in the cornucopia. The secondary reflector is designed in the same manner as the Cassegrain system. In the lowest-cost version the reflector is plane.⁶⁰ The aperture of the casshorn can be formed in a variety of ways according to the illumination concepts discussed earlier. An elliptical aperture can be designed as easily as a trapezoid or circular aperture. The feed for the casshorn is mounted at one edge of the aperture (see Figure 15-46). The problems in the design are associated with making sure that the feed pattern is not distorted by the nearby reflector surface.

A comparison of the casshorn and the cornucopia indicates that the free-space feed of the casshorn be used to provide more illumination taper for the aperture (particularly in the E plane) than the cornucopia. However, the casshorn has not proved to be superior to the cornucopia in sidelobe performance. Greater application of shielded antennas can be expected as the number of microwave installations increases.

15.6 PASSIVE REFLECTORS

Some reflector systems do not contain a feed source. These reflectors, which are called passive, usually receive and reflect a plane wave. There are three general types of passive reflectors: (1) those used in microwave links, (2) those used as radar targets, and (3) those used in a beam waveguide.

Microwave Relay

The passive reflectors⁶¹⁻⁶³ used in microwave links can be classified as repeaters or periscopes. The first reflector system controls a plane wave in the horizontal plane, and the

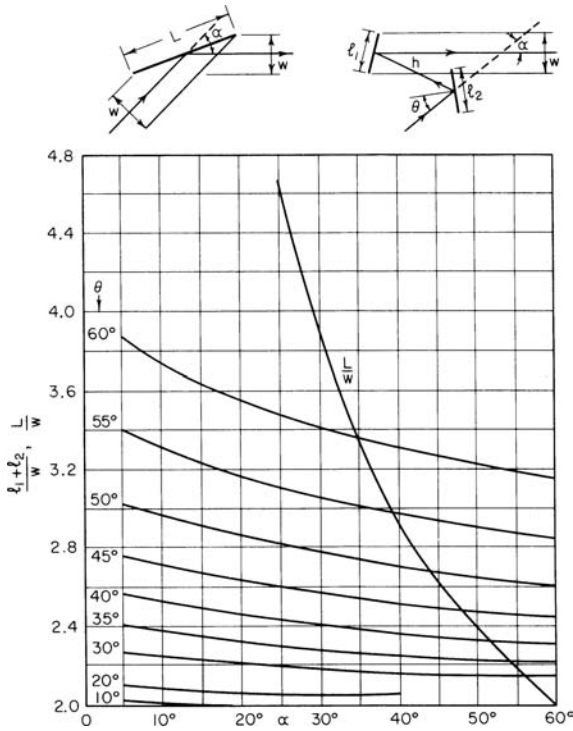


FIGURE 15-48 Variation of (normalized) mirror length with α and θ for single- and double-mirror passive repeaters

second controls the plane wave in the vertical plane. The design information for repeater reflectors given in Figure 15-48 shows the geometry of a single-reflector and a dual-reflector repeater system. The reflector surfaces should be within $\lambda/16$ of a plane. The extent of the plane surface should be minimized to reduce costs. The figure shows dual and single reflectors with the dimensions required for various angles of the ray paths. The path loss for the two systems can be calculated by assuming that the projected area seen by the transmitter and receiver is the same (and assuming that in the case of the two-reflector system all the transmitter energy captured by one reflector is reflected from the second reflector). If the ranges from the transmitter and receiver are given by d_t and d_r , then the path loss is

$$10 \log \frac{\lambda^4 d_t^2 d_r^2}{A_T A_R A}$$

where the areas are designated for transmitter, receiver, and projected area of the repeater plane surfaces.

The periscope reflector system uses a large reflector at the top of a tower and a smaller reflector on the ground. This antenna can employ a plane reflector or a curved reflector at the top of the tower. The gain obtainable from the two type surfaces is shown in Figures 15-49 and 15-50.

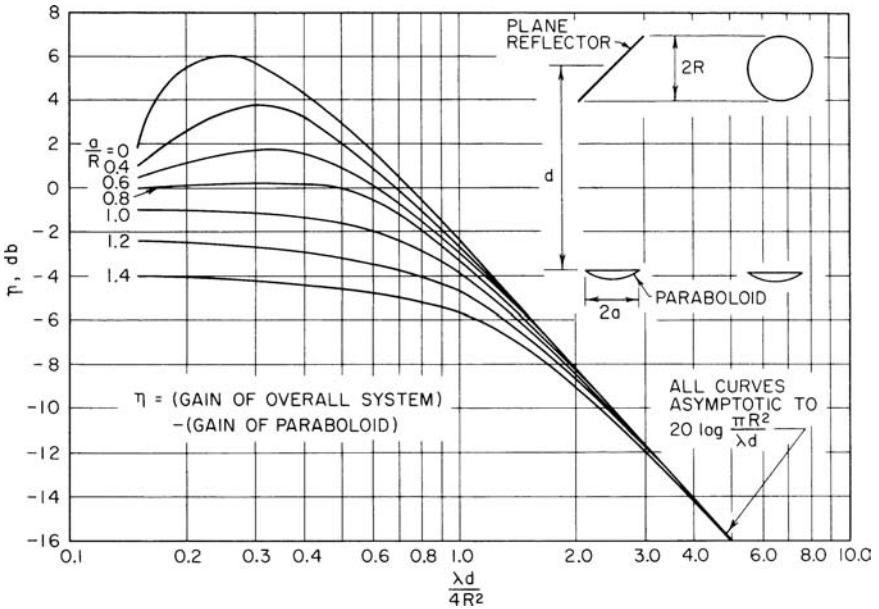


FIGURE 15-49 Relative gain of a periscope antenna system employing a plane elevated reflector

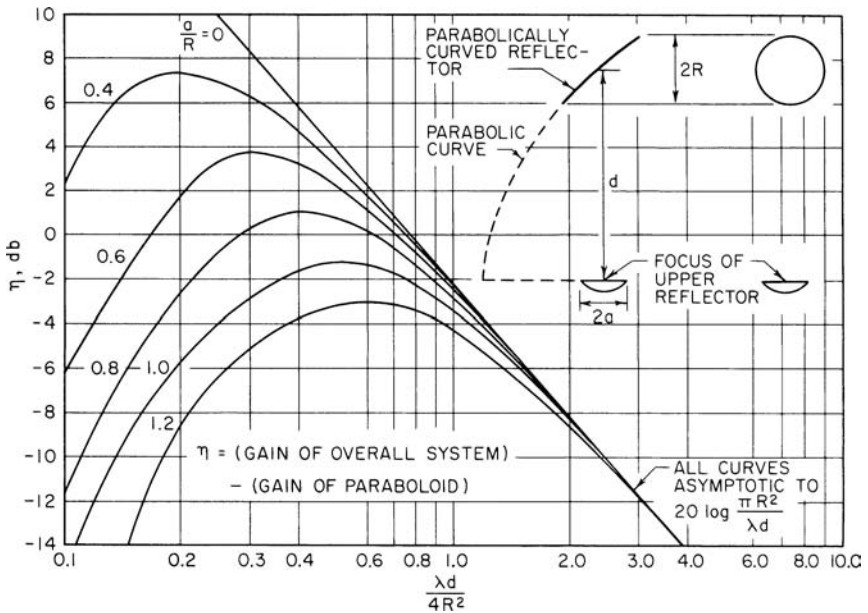


FIGURE 15-50 Relative gain of a periscope antenna system employing a curved elevated reflector

Radar Targets⁶⁴

Radar targets are passive reflectors that have a reflected-signal distribution similar to the patterns of an antenna. A number of classical targets are given in Figure 15-51. The basic problem in the design of radar targets is that of maximizing the target return.

TYPE	DIMENSIONS	MAXIMUM		ANGULAR RESPONSE	
		A_T	σ	θ	ϕ
SPHERE		$\frac{a\lambda}{2}$	πa^2	360°	360°
CYLINDER		$b\sqrt{\frac{a\lambda}{2}}$	$\frac{2\pi ab^2}{\lambda}$	360°	SHARP
FLAT PLATE		ab	$\frac{4\pi a^2 b^2}{\lambda^2}$	SHARP	SHARP
DIHEDRAL CORNER		$\sqrt{2}ab$	$\frac{8\pi a^2 b^2}{\lambda^2}$	$\pm 30^\circ$ (TO -10 DB ECHO LEVELS)	SHARP
TRIANGULAR TRIHEDRAL		$\frac{a^2}{\sqrt{3}}$	$\frac{4\pi a^4}{3\lambda^2}$	 $\phi = 0$	 $\theta = 0$
SQUARE TRIHEDRAL		$\sqrt{3}a^2$	$\frac{12\pi a^4}{\lambda^2}$		
NOTES	<p>A_T = EQUIVALENT FLAT-PLATE AREA OF TARGET σ = SCATTERING CROSS SECTION OF TARGET $\sigma = 4\pi \frac{A_T^2}{\lambda^2}$ $P_R = P_T \frac{A_R^2 A_T^2}{\lambda^4 d^4} = P_T \frac{A_R^2 \sigma}{4\pi \lambda^4 d^4}$ (FREE-SPACE TRANSMISSION) P_T = POWER EMITTED BY RADAR P_R = ECHO POWER COLLECTED BY RADAR d = DISTANCE FROM RADAR TO TARGET A_R = EFFECTIVE AREA OF RADAR ANTENNA = $\frac{G\lambda^2}{4\pi}$, WHERE G = GAIN λ = WAVELENGTH IN SAME UNITS AS d, $\sqrt{A_R}$ AND $\sqrt{A_T}$ ALL DIMENSIONS ARE ASSUMED LARGE IN WAVELENGTHS θ AND ϕ ARE ANGLES BETWEEN THE DIRECTION TO THE RADAR ANTENNA AND THE MAXIMUM RESPONSE AXIS OF THE TARGET</p>				

FIGURE 15-51 Characteristics of various types of radar targets. The columns labeled A_T and σ give the maximum values of these quantities for optimum orientation of the target. G is the gain of an antenna over an isotropic radiator.

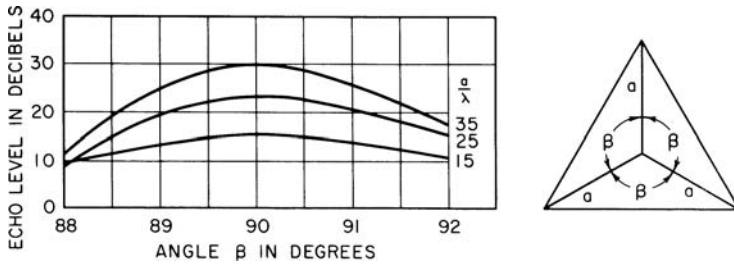


FIGURE 15-52 Effect of error in all three corner angles upon a trihedral echo response. Incident radiation is parallel to the symmetric axis.

Mechanical tolerances of the reflector surface limit the performance. For targets of the order of 20λ in extent, a $\lambda/3$ error in the edges will give a loss of 3 dB in the returned signal compared with a perfect target. Figure 15-52 shows the loss for three different corner reflectors for angular errors of the order of 2° .

The simple corner reflector is usable over an angular sector of about 30° . For a wider-coverage angle, an array of corner reflectors can be employed, or the Luneburg lens can be used with a spherical cap reflector⁶⁵ (see Figure 15-53). The figure shows that the maximum-coverage angle is of the order of 120° . Another lens that gives a theoretical return from all directions has been developed. This lens was fabricated by Emerson & Cuming, Canton, Massachusetts. The index variation is given by $n^2 = (2 - r)/r$. The dielectric constant required for small r becomes very large (at quarter radius the value is 7.5). It is possible to replace the center of the lens by a metallic sphere in order to avoid the large dielectric constant. The center portion represents only a small portion of the aperture, so that an operable system could be achieved without the expense of specialized dielectrics.

For coverage in all directions of space, a combination of 20 corner reflectors (duodecahedron) has been successfully tested.⁶⁶ The difficulty in obtaining reproducible lenses has made this reflector approach more satisfactory.

Beam Waveguide

The final use of passive reflectors is in earth stations, where typical waveguide losses cannot be tolerated. The beam-waveguide concept uses a series of large-aperture reflectors

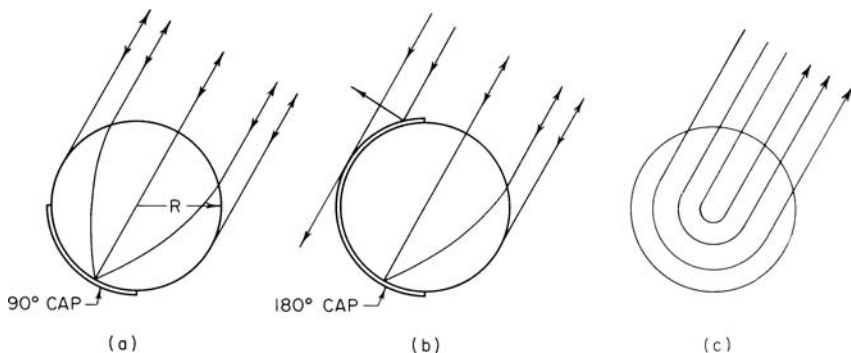


FIGURE 15-53 Lens reflectors: (a) 90° cap, (b) 180° cap, and (c) no cap

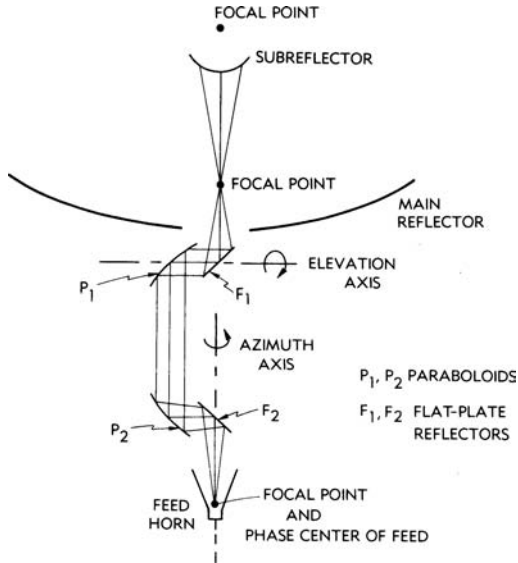


FIGURE 15-54 Geometry of a beam-waveguide feed system

to transmit energy with little loss. Initial work on such waveguides⁶⁷ used lens-type apertures. The best existing beam waveguide uses a series of paraboloids, hyperboloids, or ellipsoids. The energy is carried along the guide from one focal point to another, and at each focal point it is received from one reflector aperture and transmitted to the next reflector.

Figure 15-54 shows the geometry of the most effective beam waveguide. The reflectors used are either paraboloid-paraboloid or ellipsoid-hyperboloid.^{68,69} With the configuration chosen, it is possible to eliminate rotary joints in the transmission line. There is very little loss in beam waveguides, and they can be shielded from external signals by the use of large metallic cylinders that encase the beams as they travel between reflectors.

Polarization-Sensitive Reflectors⁷⁰

One or more reflectors in an antenna system can be made up of linear parallel reflectors that will reflect one polarization and transmit the orthogonal polarization. Such reflectors are called *transreflectors*. When they are aligned with the linear reflecting elements at 45° to the incoming polarization, half of the energy is passed and half is reflected. A second solid reflector will redirect the transmitted energy back through the linear elements with a phase that depends upon the spacing between the linear elements and the solid reflector. A spacing of one-eighth wavelength creates a 90° phase shift and circular polarization. A spacing of one-quarter wavelength creates a half-wave phase shift and a reflected polarization at 90° from the incident polarization. This system is called a *twist reflector*.

A transreflector can be used as a focusing device that has 360° coverage in the azimuth plane. The linear elements are aligned at 45° , so that half of the incident linear or circular polarization is passed and the remainder is scattered. The reflector appears as a dome that in the azimuth plane shows the near-side linear elements to be orthogonal to those on the far side. Therefore, the energy that passes through the near side of the dome is reflected at the far side.

If the surface contour matches a sphere (helisphere) or parabolic torus, the energy that reflects from the far side is focused.

The transreflector in a planar configuration can be used to provide two-feed systems in a single reflector. That is, a horizontally polarized signal can pass through vertically polarized reflecting elements. The transreflector can be placed between a paraboloidal reflector and the horizontally polarized feed at the focus. A second, vertically polarized feed can be placed at the image of the focal point in the plane reflector. This has been termed *focus splitting*.

A unique use of the transreflector and twist reflector lies in providing a thin focusing system with a very low F/D ratio.* The paraboloid with focus in face has an F/D ratio of 0.25. A transreflector mounted across the face of the curved reflector will return the energy to the curved paraboloidal surface, where it is reflected. If this surface is a twist reflector, the reflected radiation is shifted through 90° in polarization so that it now passes through the transreflector. The parallel beam of rays from the system was created by reflection at two curved surfaces (two uses of the twist-reflector surface), and therefore the surface curvature required is reduced. The F/D ratio is reduced to 0.125. A further reduction in the F/D ratio is obtained if a Cassegrain subreflector is added and the feed is placed at the vertex. The F/D ratio then is 0.08.

The transreflector or the twist reflector is made up of about ten linear elements per wavelength. Accurate planar surfaces can be easily obtained by printed-circuit card techniques. Curved surfaces are best formed with the same technology, and for this reason polarization-sensitive reflectors are most often used for physically small surfaces.

Frequency-Sensitive Reflectors^{71,31}

It is possible to design a reflector that will pass signals of one frequency band and reflect signals of other frequencies. These reflectors are called *dichroic* or frequency-sensitive surfaces (FSS). Such reflectors are made up of a large array of closely spaced resonant elements. Slots or dipoles can be used. The dipoles reflect the chosen frequency band, and the slots transmit this band. When the dipoles become small compared with a wavelength of interest, there is limited reflection and most of the energy is transmitted. When the slots become small compared with the wavelength of interest, the energy is almost totally reflected. By a proper choice of dipoles or slots, the pass-band can be at a higher or lower frequency than the second band of frequencies.

The most important constraint in the design is the spacing between the two bands of frequencies. If there is an octave band or less, the design becomes difficult and greater losses are introduced. If two octaves separate the bands, the design restrictions are not severe. For sharper response between frequency bands, it is possible to use two or more surfaces, in a manner analogous to transmission-line filter design. The remaining serious design problem is associated with polarization. The surface should preserve the incoming polarization for any angle of incidence. Most surfaces have a different response to polarization in the plane of incidence than to orthogonal polarization.

Polarization limitations have ruled out simple patch and crossed dipole elements in the frequency-sensitive array. The two most widely used elements are the *tripole* and the *Jerusalem cross* (see Figure 15-55). These elements can be more tightly packed in an array. The support of these elements presents a practical design problem since the substrate dielectric will distort the design. Thin Mylar or Kevlar is used for the array, and foam dielectric is used for support of the reflector array.

* See J. P. Shelton, U.S. Patent 4,228,437, October 14, 1980.

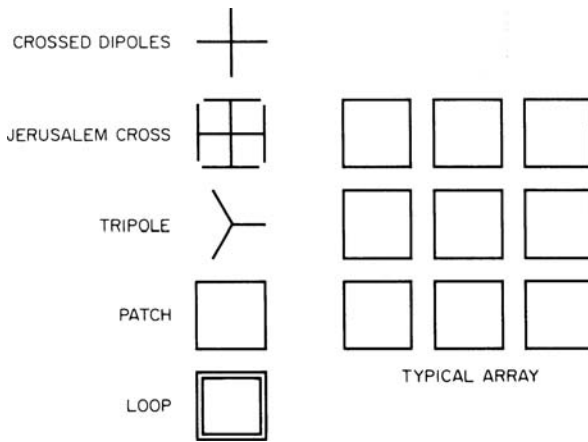


FIGURE 15-55 Elements of frequency-sensitive surfaces

15.7 LARGE-APERTURE REFLECTOR ANTENNAS

Introduction

Reflector antennas with electrically large apertures, greater than 60 wavelengths, became relatively common in the 1970s. The design of such antennas is dominated by considerations of geometrical optics,^{72,73} which are often refined by considerations of physical optics.^{74,75} There remain parameters of concern: gain, beamwidth, sidelobe level, polarization and cross polarization, antenna noise temperature, and, if more than one beam position is needed, the variation of these parameters with scan angle. The feed design or, in the case of more complex antenna optics, the design of the feed system and subreflector, intimately affect the parameters through illumination control, feed polarization properties, and spillover past the subreflector or the main reflector. Finally, the very structural design affects electrical performance through the scattering of energy from surface-panel irregularities and interpanel gaps, from subreflector and/or feed-support spars, and from the correlation of this scattering arising from the structural design and alignment procedures.

Since the birth of satellite communications, satellite-borne antennas have steadily increased in gain and complexity.⁷⁶ Higher gain derives from a need to increase EIRP in a power-limited environment.⁷⁷ Complexity may derive from several sources such as beam shaping or reconfigurability.⁷⁸ Both lead to large reflectors. Currently, reflectors electrically as large as about 200 (ACTS) and mechanically as large as 12 meters have been launched. Even larger antennas are planned for future applications. Electrically large reflectors are usually fed by simpler antennas such as horns or arrays of horns or by folded-optics systems involving subreflectors and feeds. In many instances, the antennas are designed to produce two coaxial beams with orthogonal polarizations over near-octave bandwidths. The feed systems may then terminate in complex microwave circuits involving polarizers and orthomode transducers to separate or combine orthogonally polarized beams as well as in microwave multiplexers to separate or combine different frequencies.

Electrically large reflector antennas have their principal applications in radar, where their high gain and narrow beamwidths enhance radar range and angular resolution; in radio astronomy and other deep-space applications, where the same parameters enhance sensitivity and resolution of stellar radio sources; and in microwave communications, where high

gain, low noise, and sidelobe control enhance effective radiated power relative to isotropic (EIRP), receive sensitivity, and isolation from interference. For reflector antennas of diameter greater than 60 wavelengths, gains range from 40 dBi to about 70 dBi, half-power beamwidths (HPBW) from about 1.5° to about 0.05° , and average wide-angle sidelobe levels down to -20 dBi (almost 90 dB below the peak of the main beam); dual-polarized radiation-pattern orthogonality provides isolations greater than 33 dB over the HPBWs. Of course, these properties depend upon the feed system as well as the reflectors.

Pencil-Beam Reflector Antennas

This section discusses characterization methodology for pencil-beam reflector antennas by incorporating various design options and comparing the performances of single and multi-antenna configurations.

Front-Fed Paraboloidal-Reflector Systems

The paraboloidal-reflector antenna and its design are described earlier in this chapter. Since the design is geometry-controlled and since the larger electrical apertures considered here make geometrical-optics considerations even more valid, the choices of F/D ratio, edge taper, etc., remain the same. However, the simple front-fed reflector has disadvantages due to the blocking of the aperture by the feed system and by the feed supports.^{79,80} Figure 15-56 shows two views of a front-fed paraboloidal-reflector antenna with feed supports extending

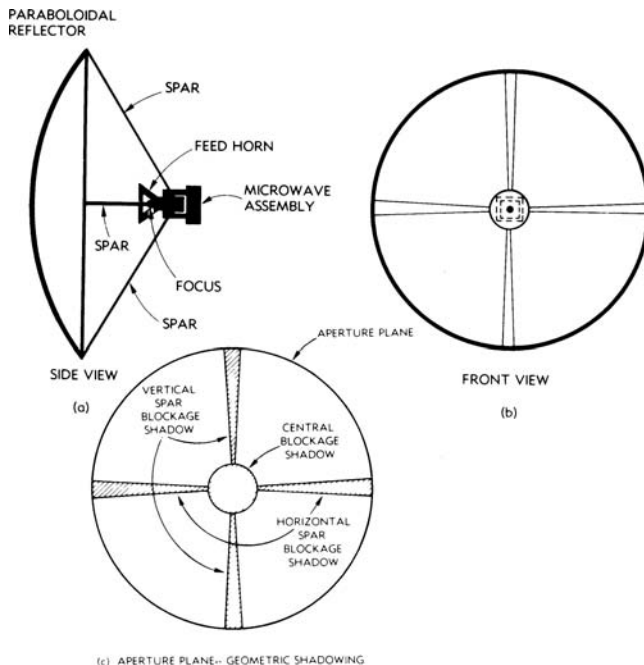


FIGURE 15-56 Front-fed paraboloidal-reflector antenna

from the reflector edge. It is seen that the feed supports intercept radiation reflected from the paraboloid. (Supports that are not on the edge intercept energy from the feed before it reaches the reflector.) In a geometrical-optics sense, the feed supports cast a shadow upon the aperture plane, shown in the third view. This shadow, called *aperture blockage*, reduces gain and affects near-in sidelobes, raising some and lowering others. In addition, the blocked radiation is scattered by the feed supports and can significantly increase wide-angle sidelobes. In general, the front-fed design with a simple feed horn has less blockage than the folded-optics designs. Additionally, the feed subtends a wider angle at the main reflector than a feed illuminating a subreflector does, and therefore the feed is a simpler, lower-gain design. However, if the feed system is complex and the microwave package abutting it is large and heavy [containing polarizers, orthomode transducers (OMTs), low-noise amplifiers (LNAs), etc.], the blocking advantage is negated somewhat. Further, the focal point is an awkward location for such electronic gear, both structurally, because the weight is not desirable at the end of a large-moment arm, and from the point of view of equipment servicing. Finally, in electrically large reflectors, in which reflector size must be minimized, the folded-optics system can be made more efficient. The front-fed reflector has an efficiency of 55 to 60 percent. Folded-optics designs can employ shaped subreflectors to increase efficiency.

Folded Optics

Pencil-beam reflectors may be illuminated by feed systems in which the path of energy from the feed (which terminates the guided-wave-waveguide assembly) is reflected by one or more subreflectors. Figure 15-57 portrays the geometry of the Cassegrain reflector system, whence it is seen that the incoming-ray trajectories are folded back in the direction of the vertex of the paraboloid P by reflection from the hyperboloidal subreflector H . The focusing properties of the system derive from the collocation of one focus of the hyperboloid H with that of the paraboloid at F_1 . An incoming plane wave from the direction of the paraboloid axis is reflected from the paraboloid and from the hyperboloid and focuses at F_2 , the other focus of the hyperboloid H . A feed with its phase center at F_2 will then receive the incident energy.

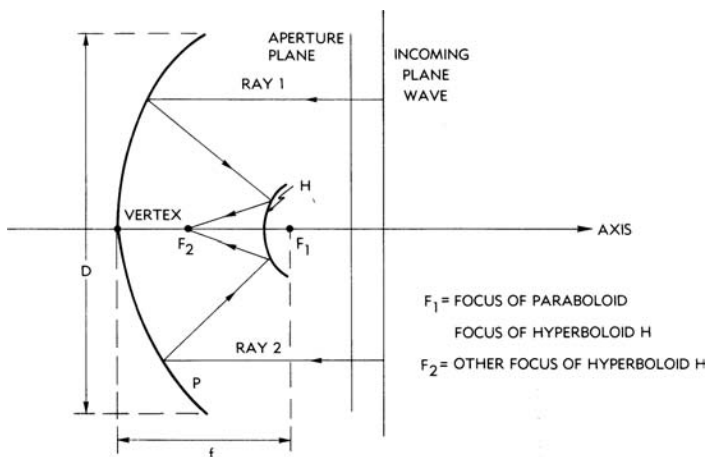


FIGURE 15-57 Cassegrain geometry

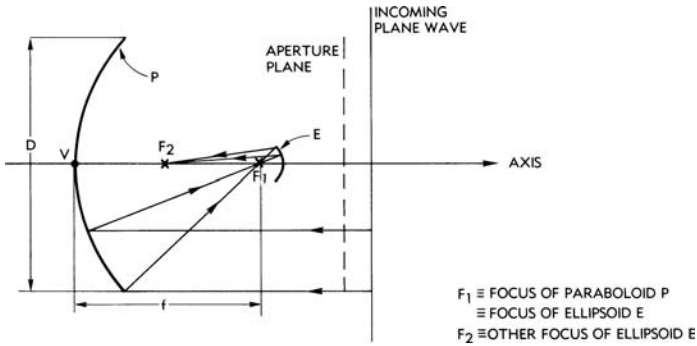


FIGURE 15-58 Gregorian geometry

Figure 15-58 portrays the geometry of the gregorian reflector system. Here the subreflector is an ellipsoid E with its near focus collocated at F_1 with the focus of the paraboloid, the ellipsoid lying on the side away from the paraboloid vertex. The phase center of the feed should then lie at F_2 , the other focus of the ellipsoid. Another configuration is shown in Figure 15-59. Here the subreflector is paraboloidal, with its focus collocated at F with the focus of the paraboloidal main reflector. The result of the reflector of the rays representing the incoming plane wave is a collimated cylinder of rays parallel to the axis. This creates the possibility of another, smaller aperture plane parallel to XX' . A feed system associated with this aperture would have a plane wavefront, or the parallel rays could be associated with an appropriate beam-waveguide system.

An immediate advantage of all three folded systems is that the feed system is now located nearer the vertex of the main reflector and, in fact, can usually be reached from access provisions at the vertex. More subtly, introduction of the additional reflection at the subreflector provides an additional degree of design freedom. Aperture distribution can be controlled to increase efficiency and decrease blockage by shaping the subreflector and the main reflector so that the surfaces deviate from the hyperboloid, ellipsoid, and paraboloid.

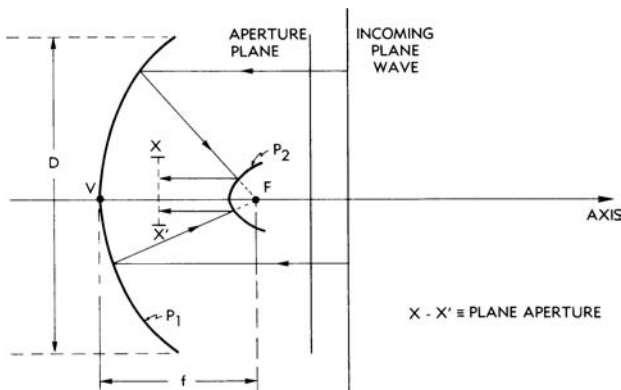


FIGURE 15-59 Planar-feed field geometry

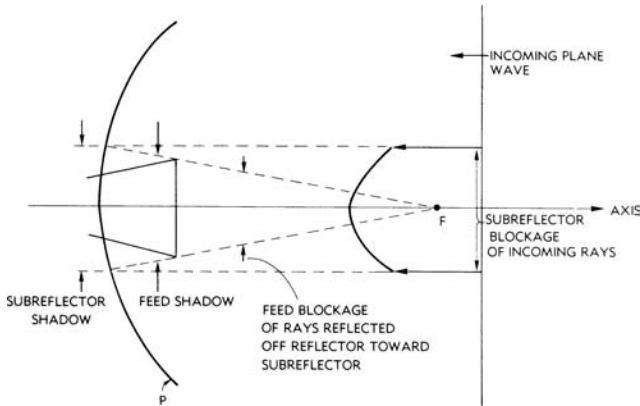


FIGURE 15-60 Blockage by feed and subreflector

The Cassegrain system is more compact than the gregorian system for the same focal length. There have been claims that gregorian systems provide improved wide-angle sidelobes in apertures of the order of 200 wavelengths. However, because of blockage of the feed system, subreflector, and supports, the sidelobes of these folded antennas are not especially low. By careful design and construction, the sidelobe envelope can be made to conform to international and Federal Communications Commission (FCC) requirements. For example, INTELSAT old Standard A antennas, which have gain in excess of 57 dBi (generally 60 dBi), require sidelobe envelopes to be less than $(29 - 25 \log \theta)$ dBi, from 1° to 48° , and less than -10 dBi beyond 48° . (θ is the pattern angle measured in a band of $\pm 1^\circ$ about the geostationary arc.)

One of the design principles used in minimizing subreflector and feed blockage is to make them equal, as shown in Figure 15-60. The subreflector blocks incoming rays, and the feed blocks rays that would hit the main reflector and be reflected to the subreflector. For the case shown, the shadows of the feed and subreflector intersect on P, the paraboloid main reflector, presenting the ideal case.

It has been found that keeping the subreflector diameter small (10 percent of the diameter of the main reflector) will keep the central-blockage sidelobes from exceeding the specifications given earlier. If this is not done, the sidelobes are high in the 1° to 10° region. The subreflector support spars tend to create sidelobe problems in a region of 5° to 20° in the plane perpendicular to the spars. It has been found useful to use a spar quadruped with all spars at 45° to the plane of concern. However, the best way to reduce sidelobes is to eliminate the blockage by the use of offset reflectors.

Sidelobes in Electrically Large Antennas

Sidelobes in a folded-optics system arise from a number of sources. These include (1) radiation-pattern sidelobes from the aperture illumination, (2) spillover and edge diffraction from the subreflector illuminated by the feed system, (3) spillover and edge diffraction from the main reflector illuminated from the subreflector, (4) central blockage (subreflector and feed system), (5) spar (subreflector support) blockage and scattering, (6) deviation of reflector surfaces from the desired contours, and (7) effects of gaps between panels.

Control of aperture illumination amplitude and phase to reduce sidelobes dates back to the earliest days of antenna theory and practice, and many references can be found in any basic antenna text. More recent techniques are discussed in two review articles.^{81,82} One approach that has become quite important is the use of offset geometry, yielding an unblocked aperture that eliminates sidelobes arising from central blockage, spar blockage, and spar scattering. This will be discussed in greater detail next.

Departure of the aperture illumination in amplitude and phase from that desired can arise from feed-aperture phase errors⁷⁹ and defocusing.⁸⁰ These will cause the radiation (diffraction) pattern to depart from the design values, but are correctible by careful feed design and positioning. Spillover and edge-diffraction effects can be avoided by illumination of the subreflector with a more deeply tapered feed pattern to about 20 dB down at the edge. For the shaped subreflector, it is possible to obtain sharp illumination taper at the main-reflector edge.^{83,84} Spillover and edge diffraction of feed radiation will give increased sidelobes in the 10° to 20° region of the pencil-beam axis.

Minimizing the central blocking and optimizing the illumination are not compatible, since the first requires smaller feed systems and the second requires larger feed systems. Equalizing the subreflector and horn blockage as discussed in the preceding subsection is the best approach. Then if the subreflector diameter can be kept to less than 10 percent of the main-reflector diameter, shaping is possible and the blockage lobes will meet the sidelobe criterion of the preceding subsection. Central blockage casts a wide beam inversely proportional to the diameter of the central blockage and thus typically ten times the beam-width and 20 dB down from the main beam. It is out of phase with the main beam and thus subtracts a fraction of 1 dB from the main beam, adds 2 to 3 dB to the first sidelobe, subtracts from the second sidelobe, etc. Spar blockage and scattering^{79,85} are not circularly symmetrical and depend upon spar shape and spar-reflector-subreflector geometry. If the spars are not mounted at the dish edges, shadows are cast on the main reflector to reduce efficiency, and blockage of the reflected energy is still present. Spar scattering effects are strongest in the forward hemisphere at an angle of $180^\circ - \xi$, where ξ is the angle between the spar axis and the aperture plane.⁷⁹ The sidelobes are strongest in the forward direction in the plane containing the spar and the beam axis. This typically occurs in the region at 5° to 15° off the main-beam axis. Spar design can minimize but not eliminate this effect. Streamlining the spar cross section is useful for the polarization perpendicular to the spar. Orienting the spars in a quadruped at 45° to the horizontal is effective for earth stations in keeping spar-induced sidelobes out of the region of the geostationary arc. Figure 15-61 shows the effects of central blockage, spar blockage, and forward (subreflector) spillover on a radiation pattern.

The remaining two sources of sidelobe energy derive from the manner in which electrically large antennas are fabricated, assembled, and installed. Most electrically large reflectors are constructed of segments (panels) assembled onto a backup structure or with the backup structure as an integral part of the panels. The subreflectors are generally single units.

The rms errors associated with fabricating a single panel or a subreflector can be controlled with care to 0.1 mm (4 mils) for a reflector or panel up to 3 m across; however, commonly used fabrication techniques permit rms errors of 0.25 to 1.27 mm (10 to 15 mils) for panels of this size. Assembly and installation errors are each of the same order, provided careful checkout procedures are used. A rule of thumb for communication and radar reflectors in the 8- to 30-m class is that the ratio of rms error to reflector diameter is 2 to 5 times 10^{-5} . Some electrically large antennas have been built to better tolerances for radio astronomy, but they usually require special protection from the environment. The pattern effects of the errors depend upon their statistical nature.^{79,86-89}

It has been shown⁹⁰ that when error correlation intervals are small, the results obtained by Ruze^{86,87} for axial gain apply (see Figure 15-62). However, because panel-fabrication error contributions tend to be of the same order or smaller than assembly and installation

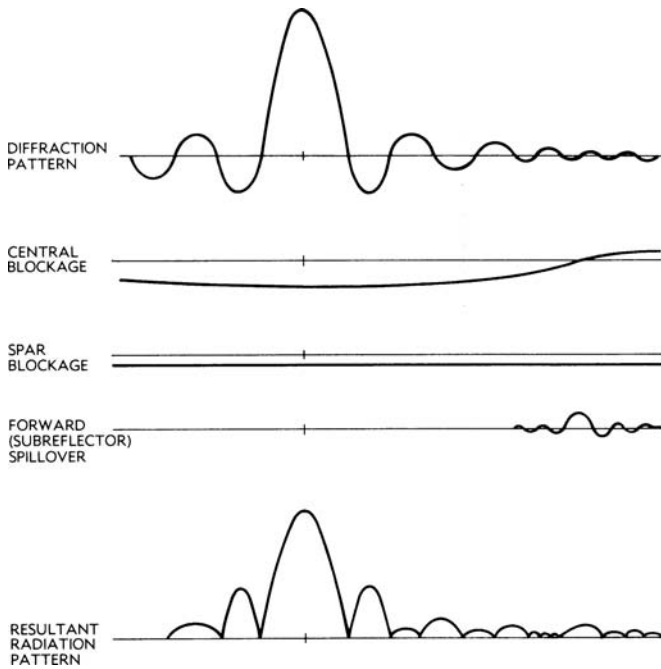


FIGURE 15-61 Blockage effects on sidelobe structure

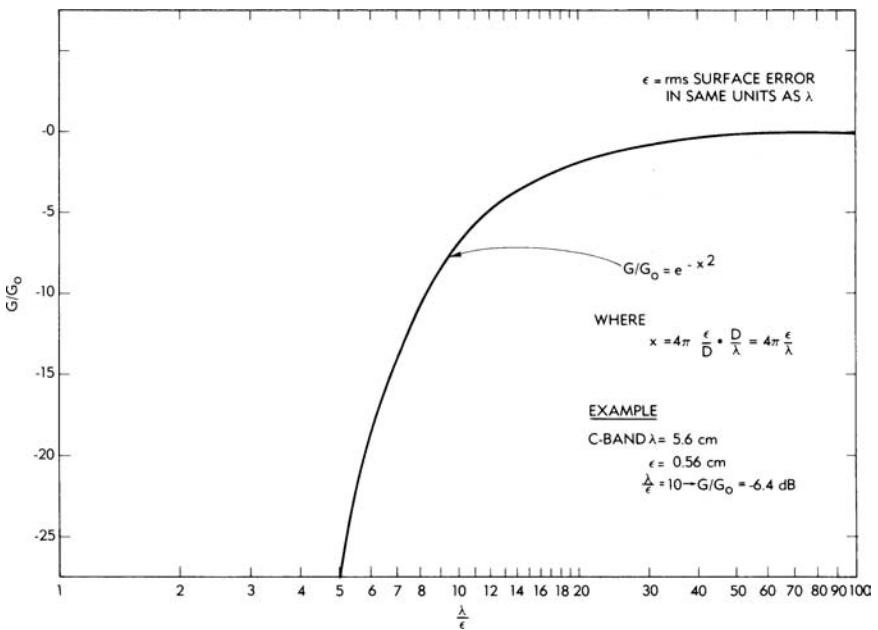


FIGURE 15-62 Normalized loss due to finite surface error

error contributions, the large correlation errors of the assembly can dominate. Therefore, you cannot reasonably assume small correlation intervals in most large antennas. Bao⁹⁰ has shown that the effect of these large correlation intervals is to reduce the gain loss and that the sidelobes are significantly lower than predicted under Ruze's assumptions regarding correlation. In a similar fashion the distortions due to wind and gravity have long correlation intervals and hence less effect than predicted by the small-correlation-interval theory. The large correlation intervals increase the sidelobes in the region near the main beam, while the small correlation intervals (random errors) contribute significantly to wide-angle sidelobes. Papers by Rahmat-Samii et al.⁹¹⁻⁹⁴ and a review paper by Corkish⁹⁵ on the effects of reflector surface distortions on sidelobe levels cover the various approaches to this subject in some detail and provide a wealth of references.

An excellent treatment on very large earth station antenna sidelobes was given by Korvin and Kreutel,⁸⁰ who presented results of both theoretical and measured studies for apertures ranging from 100 to 600 wavelengths. Figure 15-63 presents some of these results.

As real-time control techniques and data processing have become more capable in the past decade, the possibility of correcting for distortions has become a reality. Real-time adaptive optics for optical telescopes was demonstrated in 1989, mainly to remove distortion due to the atmosphere.⁹⁶ Adaptive phased-array feeds can be used in a similar manner to compensate for deterioration of scanning.⁹⁷ Schell's multiple antenna,⁹⁸ while invented for other purposes, embodies all the necessary characteristics for adaptive removal of both atmospheric and surface distortions. The advent of modern control electronics and drive mechanisms should make the multiple antenna a candidate approach for large-aperture reflectors.

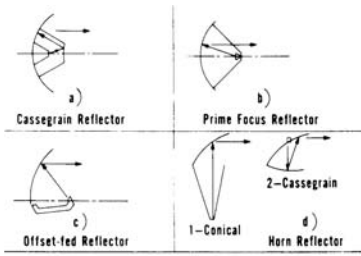
The increased interest in large, spacecraft-borne antennas and in sidelobe control has given rise to increased interest in and use of offset configurations, especially for multibeam and scanning applications.^{99,100} The conventional pencil-beam Cassegrain and gregorian configurations have been studied to provide design techniques for the offset cases.¹⁰¹⁻¹⁰³ A technique for increasing antenna efficiency by extending the subreflector in an offset Cassegrain antenna has been analyzed.¹⁰⁴ When offset Cassegrain or gregorian reflector systems are used in beam-scanning or multibeam applications, feeds must be moved or located in the positions that provide optimal beam performance. This turns out to be a hyperboloid for the offset Cassegrain^{105,106} and an ellipsoid for the offset Gregorian.¹⁰⁷

High-Efficiency Dual-Reflector Systems

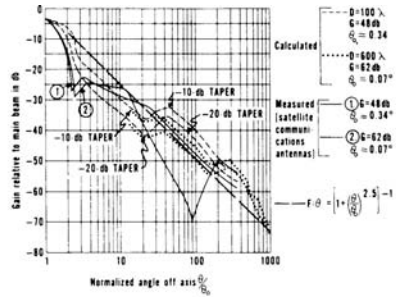
Electrically large pencil-beam antennas are costly. Cost is almost a linear function of aperture area (reflector surface). The higher the efficiency, the smaller the aperture required to collect incoming waves or to provide gain for a transmitted wave. Antenna efficiency is the product of a number of factors, each of which represents the degree with which the energy is protected against diversion from the desired direction or polarization by a particular mechanism. The factors include effects of (1) aperture illumination (η_i), (2) spillover around the main reflector and subreflector (η_{s1} , η_{s2}), (3) blockage (η_b), (4) reflector-surface deviations (η_r), (5) polarization purity (η_p), and (6) losses in waveguide and feed (η_0). These efficiency factors are defined in Table 15-1.

Spillover, blockage, and surface tolerance effects have been discussed earlier in terms of sidelobe control. Improvement in efficiency and reduction of sidelobe effects go hand in hand for these factors. Polarization purity is controlled by feed-system polarization properties, by curvature of the reflector, by symmetry, by scattering properties of the blockage elements, and, for circular polarization, by internal reflections in the waveguide and feed system.

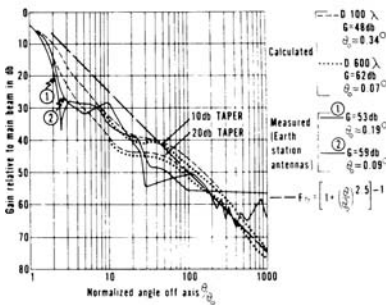
Control of aperture illumination¹⁰⁸⁻¹¹⁰ is a major factor in improving efficiency. Techniques for controlling illumination can also minimize the effects of central blockage,



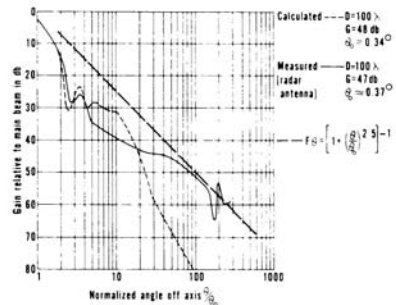
BASIC CONFIGURATIONS OF ANTENNAS



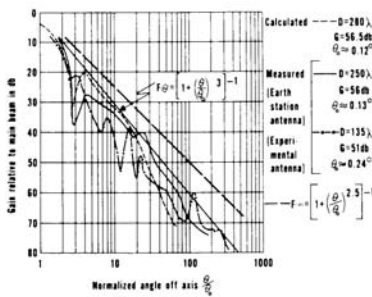
CASSEGRAIN-FED REFLECTOR RADIATION DIAGRAMS



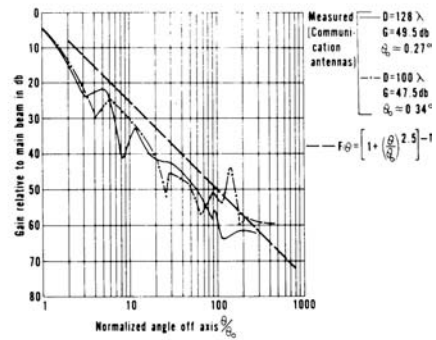
PRIME-FOCUS-FED PARABOLIC REFLECTOR



OFFSET-FED PARABOLIC REFLECTOR



HORN REFLECTOR (CONICAL) ANTENNA



HORN REFLECTOR (CASSEGRAIN) ANTENNA

FIGURE 15-63 Practical sidelobe-level envelope

as will be seen. If maximum gain is desired, the object is to shape the subreflector and the main reflector so that the aperture illumination is as nearly uniform as possible, with minimum spillover and with minimum energy blocked.

Three basic conditions are used to shape the two reflectors: (1) Path length from phase center of feed to aperture plane is constant for all paths. For the j th ray, as shown in Figure 15-64,

TABLE 15-1 Antenna Efficiency Factors

$$\eta_i = \frac{1}{A} \left| \int F dA \right|^2 / \left| \int F \right|^2 dA \quad (F \text{ is illumination over aperture } A.)$$

$$\eta_{s1} = \text{power on main reflector/(power from subreflector—scattered power)}$$

$$\eta_{s2} = \text{power on subreflector/power from feed system}$$

$$\eta_b = \text{power blocked/total power from main reflector}$$

$$\eta_r = \exp [-(4\pi/\lambda)^2] \quad (= \text{rms error. This is worst-case approximation.})$$

$$\eta_0 = \text{power radiated by feed/power input to feed system}$$

$$\eta_p = \text{aperture power of desired polarization/total aperture power}$$

$(\overline{F_2 S_j} + \overline{S_j M_j} + \overline{M_j A_j}) = \text{path length from feedpoint to aperture plane} = r_j + z_j + (r_j \cos \theta_j - l + z_j) \cos \theta_j = \text{constant.}$ (2) Aperture illumination is constant. It can be shown^{109,110} that this requirement leads to

$$x_j^2/x_{\max}^2 = \left(\int_0^{\theta_j} E_f^2 \sin \theta d\theta \right) / \left(\int_0^{\theta_{\max}} E_f^2 \sin \theta d\theta \right) \quad (15-40)$$

where $E_f = E_f(\theta)$ is the circularly symmetric pattern of the feed horn. (3) Snell's law must apply. This leads to the expressions

$$\frac{1}{r} \frac{dr}{d\theta} = \tan\left(\frac{\theta + \psi}{2}\right) \quad -\frac{dz}{dx} = \tan \frac{\psi}{2} \quad (15-41)$$

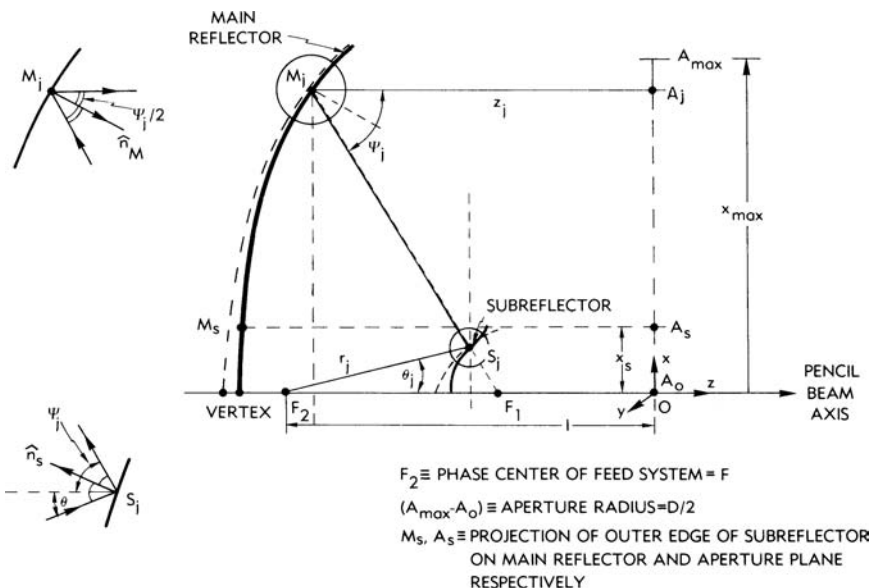


FIGURE 15-64 Geometrical optics (GO) shaping of main reflector and subreflectors

These three conditions (four equations) can be solved on a computer to yield the two surfaces desired. But the general nature of the solution can be deduced from the physical problem. It is desirable to make the aperture illumination more uniform when it is too high in the center. If the slope of the subreflector near the center were altered to reflect central rays from the feed away from the center of the main reflector, this would have the desired effect. The dotted curve in front of the subreflector represents such a distortion. The main reflector must now be reshaped to maintain the focused beam. This results in the dotted curve behind the main reflector.

If the process is exaggerated near the middle of the subreflector, a hole would result in the illumination of the main reflector. The energy here is normally blocked by the subreflector so that there is no basic change in the radiation pattern near the main beam. There will be a reduction in the energy scattered by the subreflector and, therefore, lower scattered sidelobes at wider angles. It has been indicated¹¹¹ that you can optimize the shaping of dual-reflector systems for G/T_{antenna} , rather than G alone. Since T_{antenna} has a major component arising from sidelobes intercepting the earth, this entails sidelobe control as well as gain optimization.

At the edge of the subreflector the surface can be shaped to direct energy inward and so give a shorter and deeper illumination taper at the edge of the main reflector to reduce spillover. The shaping of the subreflector edges can compensate for deeper tapering of the energy incident upon this surface. Therefore, a system with less spillover past the subreflector and with under-illumination of the main reflector can use the shaping of the subreflector edges to compensate for the under-illumination. Thus, it is seen that reflector and subreflector shaping not only can improve the aperture illumination but can also reduce spillover past both reflectors.

It has been found that the simple geometrical-optics analysis of shaping is not sufficiently accurate.¹¹² Improved performance can be obtained by considering diffraction effects at the subreflector.⁸⁴ The system improvement has been limited in bandwidth, and a third approach⁸⁴ modifies geometrical-optics design to give good performance over a wider bandwidth.

With the growing popularity of offset configurations, shaping theory has been applied to improve the efficiency of offset-fed dual reflectors. The geometric-optics approach developed by Galindo¹⁰⁸ has been particularized for the Cassegrain and gregorian configurations^{113–115} and extended to show that the geometric-optics–based design produces very similar results when geometric-optics–based predictions are compared with those of physical optics and the geometric theory of diffraction in the aperture.¹¹⁶

15.8 DIFFRACTION SHAPING

To overcome the limitations of GO shaping, several methods have been attempted. In an early work,⁸⁴ reflector shaping by diffraction synthesis was carried out using the method of spherical wave expansion (SWE). The more general framework of “reflector surface expansion coefficient optimization” was effectively applied to synthesizing the aperture field using the GO method¹¹⁷ and later employed in methods that incorporate GO algorithm and physical optics (PO) analysis such as Nomoto and Watanabe¹¹⁸ and Bergmann et al¹¹⁹ (the *indirect* method). This concept was further extended to methods that carry out full PO synthesis (shaping) of reflector antennas with circular aperture and single feed, such as the single-reflector contoured beam antennas in Bergmann et al¹¹⁹ (the *direct* method), and the dual-reflector pencil-beam antennas in Nomoto and Watanabe¹¹⁸ and Schlobohm and Arndt.¹²⁰ The expansion functions used in these works, such as the Fourier series and polynomials, are usually heuristically chosen. In the PO diffraction synthesis (shaping) method introduced in Rahmat-Samii and Mumford,¹²¹ a set of orthogonal global expansion functions was applied effectively to represent reflectors with circular apertures. This orthogonal representation will be used and generalized in the diffraction synthesis technique presented in this chapter.

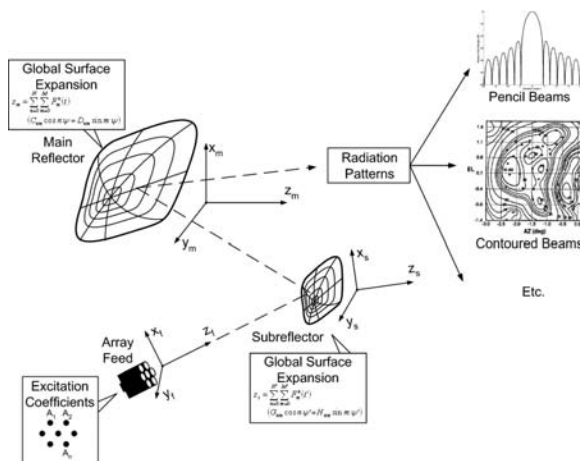


FIGURE 15-65 An overview of the generalized diffraction synthesis technique. The unknown coefficients (C_{nm} , D_{nm} , G_{nm} , H_{nm} , and A_n), which characterize the antenna system, are to be determined using optimization algorithms.

A generalized diffraction synthesis technique that can be used to synthesize single- and dual-reflector antennas with a single feed or an array feed is shown in Figure 15-65. There are several features that distinguish this technique from others. First, the previous notion of “reflector surface expansion and coefficients optimization” has been generalized to “antenna system characterization and parameters optimization,” in which the reflector surface expansion becomes a portion of the characterization process, and the expansion coefficients consist of a subset of the antenna parameters to be optimized. This conceptual generalization allows us to solve the problems of (1) reflector shaping with general array feed and (2) simultaneous synthesis of reflectors and feeds, which are not addressed in the PO synthesis methods published so far.

Another feature that makes the current technique powerful is that the global surface expansion¹²¹ is used and generalized to reflector aperture boundaries that may be circular (elliptical), square (rectangular), or any intermediate rounded-corner shapes described by the superquadric functions.¹²² The shaped reflectors are characterized by having smooth surfaces, well-defined circumferences, and continuous surface derivatives.

Also unique in the method is that the PO/physical theory of diffraction (PTD) technique is used to analyze the antennas. The PTD fringe field^{123,124} is particularly useful when PO analysis is not sufficiently accurate for very stringent requirements on, for example, sidelobe levels at far-angular regions and cross-polarization control. The edge diffraction and near-field effects are included automatically in the course of diffraction synthesis.

The generalized diffraction synthesis technique can be applied to produce various radiation patterns such as pencil-beams and contoured beams, and to improve the radiation characteristics of an existing canonical or distorted system. It is believed that the scope of the application of this technique embraces the majority of commonly used reflector antenna configurations.¹²⁵

REFERENCES

1. H. Jasik, *Antenna Engineering Handbook* (New York: McGraw-Hill, 1961).
2. Y. Rahmat-Samii, “Reflector Antennas,” *Antenna Handbook*, Y. T. Lo and S. W. Lee (eds.), Chap. 15 (New York: Van Nostrand Reinhold, 1988): 15-1–15-124.

3. A. W. Love (ed.), *Reflector Antennas* (New York: IEEE Press, 1978).
4. E. C. Collin and F. J. Zucker (eds.), *Antenna Theory*, Parts 1 and 2 (New York: McGraw-Hill, 1969).
5. S. Silver (ed.), *Microwave Antenna Theory and Design*, Chap. 12 (New York: McGraw-Hill, 1949).
6. P. J. B. Clarricoats and G. T. Poulton, "High-Efficiency Microwave Reflector Antennas—A Review," *Proc. IEEE*, vol. 65, no. 10 (October 1977): 1470–1504.
7. A. W. Love, *Electromagnetic Horn Antennas* (New York: IEEE Press, 1976).
8. A. W. Rudge et al (eds.), *The Handbook of Antenna Design*, Volumes I and II, IEE Electromagnetic Wave Series 15 (London: Peter Peregrinus Ltd., 1982).
9. M. I. Skolnik (ed.), *Radar Handbook*, Chap. 9 (New York: McGraw-Hill, 1970).
10. P. J. B. Clarricoats, "Feeds for Reflector Antennas—A Review," Part 1: Antennas, *Second Intl. Conf. Antennas Propag.* (York, England, April 13–16, 1981): 309–317.
11. A. W. Rudge and N. A. Adatia, "Offset-Parabolic-Reflector Antennas: A Review," *Proc. IEEE*, vol. 66, no. 12 (December 1978): 1592–1618.
12. Y. Rahmat-Samii, "Useful Coordinate Transformations for Antenna Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-27 (July 1979): 571–574.
13. D. W. Duan and Y. Rahmat-Samii, "Novel Coordinate System and Rotation Transformations for Antenna Applications," *Electromagnetics*, 15 (January 1995): 17–40.
14. V. Galindo-Israel and Y. Rahmat-Samii, "A New Look at a Fresnel Field Computation Using the Jacobi-Bessel Series," *IEEE Trans. Antennas Propagat.*, vol. AP-29 (November 1981): 885–898.
15. A. C. Ludwig, "Calculation of Scattered Patterns from Symmetrical Reflectors," Tech. Rep. No. 32–1430, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, February 1970.
16. Y. Rahmat-Samii and V. Galindo-Israel, "Shaped Reflector Antenna Analysis Using the Jacobi-Bessel Series," *IEEE Trans. Antennas Propagat.*, vol. AP-28, no. 4 (July 1980): 425–435.
17. V. Galindo-Israel and R. Mittra, "A New Series Representation for the Radiation Integral with Application to Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (September 1977): 631–641.
18. Y. Rahmat-Samii, R. Mittra, and V. Galindo-Israel, "Computation of Fresnel and Fraunhofer Fields of Planar Apertures and Reflector Antennas by Jacobi-Bessel Series—A Review," *J. Electromagnetics*, vol. 1, no. 2 (April–June 1981): 155–185.
19. O. M. Bucci, G. Franceschetti, and G. D. Elia, "Fast Analysis of Large Antennas—A New Computational Philosophy," *IEEE Trans. Antennas Propagat.*, vol. AP-28 (May 1980): 306–310.
20. Y. Rahmat-Samii and R. Chueng, "Nonuniform Sampling Techniques for Antenna Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-35 (March 1987): 268–279.
21. J. B. Keller, "Geometrical Theory of Diffraction," *J. Opt. Soc. Am.*, vol. 52 (1962): 116–130.
22. S. W. Lee, P. Cramer, Jr., K. Woo, and Y. Rahmat-Samii, "Diffraction by an Arbitrary Subreflector: GTD Solution," *IEEE Trans. Antennas Propagat.*, vol. AP-27 (May 1979): 305–316.
23. R. G. Kouyoumjian, "The Geometrical Theory of Diffraction and Its Application," *Topics in Applied Physics, Volume 3: Numerical and Asymptotic Techniques in Electromagnetics* (Heidelberg: Springer-Verlag, 1975): 165–215.
24. Y. Rahmat-Samii, P. O. Iversen, and D. W. Duan, "GTD, PO, PTD and Gaussian Beam Diffraction Analysis Techniques Applied to Reflector Antennas," *Applied Computational Electromagnetic Journal (ACES)*, vol. 6, no. 1 (Summer 1991): 6–30.
25. S. W. Lee and G. A. Deschamps, "A Uniform Asymptotic Theory of Electromagnetic Diffraction by a Curved Wedge," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 1 (January 1976): 25–34.
26. R. G. Kouyoumjian and P. H. Pathak, "A Uniform Asymptotic Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *Proc. IEEE*, vol. 62, no. 11 (November 1974): 1448–1461.
27. Y. Rahmat-Samii and R. Mittra, "Spectral Analysis of High-Frequency Diffraction of an Arbitrary Incident Field by a Half-Plane—Comparison with Four Asymptotic Techniques," *Radio Sci.*, vol. 13 (January–February 1978): 31–48.

28. J. Boersma and Y. Rahmat-Samii, "Comparison of Two Leading Uniform Theories of Edge Diffraction with the Exact Uniform Asymptotic Solution," *Radio Sci.*, vol. 15 (November–December 1980): 1179–1194.
29. Y. Rahmat-Samii and S. W. Lee, "Vector Diffraction Analysis of Reflector Antennas with Mesh Surfaces," *IEEE Trans. Antennas Propagat.*, vol. AP-33 (January 1985): 76–90.
30. W. Imbriale, V. Galindo-Israel, and Y. Rahmat-Samii, "On the Reflectivity of Complex Mesh Surfaces," *IEEE Trans. Antennas Propagat.*, vol. AP-39, no. 7 (September 1991): 1352–1365.
31. Y. Rahmat-Samii and A. N. Tulintseff, "Diffraction Analysis of Frequency Selective Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 4 (April 1993): 476–487.
32. A. Ip and Y. Rahmat-Samii, "Analysis and Characterization of Multilayered Reflector Antennas: Rain/Snow Accumulation and Deployable Membrane," *IEEE Trans. Antennas Propagat.*, vol. 46, no. 11 (November 1998): 1593–1605.
33. A. Miura and Y. Rahmat-Samii, "RF Characteristics of Spaceborne Antenna Mesh Reflecting Surfaces: Application of Periodic Method of Moments," *Microwave and Optical Technology Letters*, vol. 47, issue 4 (November 2005): 365–370.
34. E. F. Harris, "An Experimental Investigation of the Corner-Reflector Antenna," *IRE Proc.*, vol. 41 (May 1953): 645.
35. N. Inagaki, "Three Dimensional Corner Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (July 1974): 580.
36. J. L. Wong and H. E. King, "A Wideband Corner Reflector Antenna for 240–400 mHz," *IEEE Antennas Propagat. Symp. Dig.*, Quebec, 1980, p. 320.
37. R. C. Spencer et al, "Double Parabolic Cylinder Pencil Beam Antenna," *IRE Conv. Rec.*, part 1, *Antennas and Propagation*, 1954.
38. Y. Rahmat-Samii, J. Huang, B. Lopez, M. Lou, E. Im, S. Durden, and K. Bahadori, "Advanced Precipitation Radar Antenna: Array-Fed Offset Membrane Cylindrical Reflector Antenna," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 8 (August 2005): 2503–2515.
39. N. N. Fullilove et al, "The Hourglass Scanner," *IRE Conv. Rec.*, Part I (1959): 190.
40. R. C. Benoit and W. M. Furlow, "Wullenweber-Type UHF Radio DF," *IRE Conv. Rec.*, 1955.
41. K. S. Kelleher, "A New Wide Angle Microwave Reflector," *Tele-Tech & Electronic Ind.*, (June 1953).
42. R. A. Hoferer and Y. Rahmat-Samii, "RF Characterization of an Inflatable Parabolic Torus Reflector Antenna for Space-Borne Applications," *IEEE Trans. Antennas Propagat.*, vol. 46, no. 10 (October 1998): 1449–1457.
43. E. G. Njoku, Y. Rahmat-Samii, J. Sercel, W. J. Wilson, and M. Moghaddam, "Evaluation of an Inflatable Antenna Concept for Microwave Sensing of Soil Moisture and Ocean Salinity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no.1 (January 1999): 63–78.
44. Y. Rahmat-Samii and S. W. Lee, "Directivity of Planar Array Feeds for Satellite Reflector Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-31 (May 1983): 463–470.
45. Y. Rahmat-Samii, "A Comparison Between GO/Aperture Field and Physical Optics Methods for Offset Reflectors," *IEEE Trans. Antennas Propagat.*, vol. AP-32 (March 1984): 301–306.
46. T. S. Chu and R. H. Turrin, "Depolarization Properties of Offset Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (May 1973): 339–345.
47. A. C. Ludwig, "The Definition of Cross Polarization," *IEEE Trans. Antennas Propagat.*, vol. AP-21, no. 1 (January 1973): 116–119.
48. P. G. Ingerson and W. V. T. Rusch, "Radiation from a Paraboloid with an Axially Defocused Feed," *IEEE Trans. Antennas Propagat.*, vol. AP-21, no. 1 (January 1973): 104–106.
49. Y. T. Lo, "On the Beam Deviation Factor of a Parabolic Reflector," *IRE Trans. Antennas Propagat.*, vol. AP-8 (May 1960): 347–349.
50. J. Ruze, "Lateral-Feed Displacement in a Paraboloid," *IEEE Trans. Antennas Propagat.*, vol. AP-13 (September 1965): 660–665.
51. W. A. Imbriale, P. G. Ingerson, and W. C. Wong, "Large Lateral Feed Displacements in a Parabolic Reflector," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (November 1974): 742–745.

52. K. S. Kelleher, "Relations Concerning Wavefronts and Reflectors," *J. App. Phys.*, vol. 21, no. 6 (June 1950): 573.
53. H. J. Riblet and C. B. Barker, "A General Divergence Formula," *J. App. Phys.*, vol. 19, no. 1 (January 1948): 63.
54. T. Li, "A Study of Spherical Reflectors as Wide-Angle Scanning Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-7 (1959): 223.
55. G. Stavits and A. Dorne, "Horns and Reflectors," *Very High Frequency Techniques*, vol. 1, Chap. 6 (New York: McGraw-Hill, 1947).
56. A. C. Ludwig, "Conical Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (March 1972): 146.
57. A. S. Dunbar, "Calculations of Doubly-Curved Reflectors for Shaped Beams," *IRE Proc.*, vol. 3b (October 1948): 1289.
58. F. Briskell and B. S. Wescott, "Reflector Design as an Initial-Value Problem," *IEEE Trans. Antennas Propagat.*, vol. AP-24 (July 1976): 531.
59. S. R. Jones and K. S. Kelleher, "A New Low-Noise High Gain Antenna," *IEEE Conv. Rec.*, part I, 1963.
60. K. S. Kelleher, U.S. Patent 3,283,331.
61. W. C. Jakes, Jr., "A Theoretical Study of an Antenna-Reflector Problem," *IRE Proc.*, vol. 41 (February 1953): 272.
62. R. E. Greenquist and A. J. Orlando, "An Analysis of Passive Reflector Antenna Systems," *IRE Proc.*, vol. 42 (July 1954): 1173.
63. E. Bedrosian, "The Curved Passive Reflector," *IRE Trans. Antennas Propagat.*, vol. AP-3 (October 1955): 168.
64. S. D. Robertson, "Targets for Microwave Radar Navigation," *Bell Syst. Tech. J.*, vol. 26 (October 1947): 852.
65. K. S. Kelleher, U.S. Patent 2,866,971.
66. R. Berg, U.S. Army MERADCOM, Fort Belvoir, Maryland, private communication. Note C. H. Lockwood, U.S. Patent 3,039,093.
67. G. Goubau and F. Schwering, "On the General Propagation of Electromagnetic Wave Beams," *IRE Trans. Antennas Propagat.*, vol. AP-9 (May 1961): 248.
68. M. Mizusawa and T. Kitsuregawa, "A Beamwaveguide Feed Having a Symmetric Beam for Cassegrain Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (November 1973): 884.
69. B. Claydon, "Beam Waveguide Feed for a Satellite Earth Station Antenna," *Marconi Rev.*, vol. XXXIX, no. 201 (second quarter 1976): 81-116.
70. L. K. DeSize and J. F. Ramsay, "Reflecting Systems," *Microwave Scanning Antennas*, R. C. Hansen (ed.), vol. 1, Chap. 2 (New York: Academic Press, 1964).
71. V. D. Agrawal and W. A. Imbriale, "Design of a Dichroic Cassegrain Subreflector," *IEEE Trans. Antennas Propagat.*, vol. AP-27, no. 4 (July 1979).
72. F. S. Holt, "Application of Geometric Optics to Design and Analysis of Microwave Antennas," Air Force Cambridge Res. Lab. Tech. Rep. AFCRL-67-0501, Bedford, MA, September 1967.
73. M. Kline and I. W. Kay, *Electromagnetic Theory and Geometric Optics*, Interscience Pure and Applied Mathematics ser., vol. XII (New York: Interscience Publishers, a division of John Wiley & Sons, 1965).
74. W. V. T. Rusch, "A Comparison of Geometrical and Integral Fields from High Frequency Reflectors," *IEEE Proc.*, vol. 62, no. 11 (November 1974): 1603-1604.
75. W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas* (New York: Academic Press, 1970).
76. A. J. Zaghoul et al, "Advances in Multibeam Communications Satellite Antennas," *Proc. IEEE*, vol. 78, no. 7 (July 1990): 1214-1232.
77. F. M. Naderi and S. J. Campanella, "NASA's Advanced Communications Technology Satellite (ACTS): An Overview of the Satellite, the Network, and the Underlying Technologies," AIAA 12th International Conference on Satellite Communications, Arlington, VA, March 1988.

78. R. R. Persinger, S. O. Love, M. F. Caulfield, and A. I. Zaghoul, "INTELSAT VI Antenna System," *COMSAT Tech. Rev.*, vol. 21, no. 1 (Spring 1991).
79. A. D. Monk, "The Prediction and Control of the Wide-Angle Sidelobes of Satellite Earth Station Antennas," *Marconi Rev.*, vol. XLIII, no. 217 (second quarter 1980): 74–95.
80. W. Korvin and R. W. Kreutel, "Earth Station Radiation Diagram with Respect to Interference Isolation Capability: A Comparative Evaluation," *AIAA Progress in Astronautics and Aeronautics: Communications Satellites for the 70's: Technology*, vol. 25 (Cambridge, MA: The M.I.T. Press, 1971): 535–548.
81. H. E. Schrank, "Low Sidelobe Reflector Antennas," *IEEE AP-S Newsletter* (April 1985): 5–15.
82. C. J. Sletten, W. Payne, and W. Shillue, "Offset Dual Antennas for Very Low Sidelobes," *Microwave J.* (May 1986): 221–240.
83. W. F. Williams, "High Efficiency Antenna Reflector," *Microwave J.*, vol. 8, no. 7 (July 1965): 79–82.
84. P. J. Wood, "Reflector Profiles for the Pencil-Beam Cassegrain Antenna," *Marconi Rev.*, vol. XXXV, no. 185 (second quarter 1972): 121–138.
85. G. Hyde and R. W. Kreutel, "Earth Station Antenna Sidelobe Analysis," *IEEE ICC Conf. Rec.* (1969): 32–7–32–14.
86. J. Ruze, "Physical Limitations on Antennas," Ph.D. thesis, MIT Res. Lab. Electron. Tech. Rep. 248, October 1952.
87. J. Ruze, "The Effect of Aperture Errors on the Antenna Radiation Pattern," *Supplemento al Nuovo Cimento*, vol. 9, no. 3 (1952): 364–380.
88. J. Robieux, "Influence de la Precision de Fabrication d'une Antenne sur Ses Performances," *Annales de radioélectricité*, vol. XI, no. 43 (January 1956): 29–56.
89. H. Zucker, "Gain of Antennas with Random Surface Deviations," *Bell Syst. Tech. J.*, vol. 47 (1968): 1637–1651.
90. V. T. Bao, "Influence of Correlation Interval and Illumination Taper in Antenna Tolerance Theory," *IEE Proc.*, vol. 116, no. 2 (February 1969): 195–202.
91. Y. Rahmat-Samii, "Effects of Deterministic Surface Distortions on Reflector Antenna Performance," *Annales des Telecommunications*, France, vol. 40 (August 1985): 350–360.
92. H. Ling, Y. T. Lo, and Y. Rahmat-Samii, "Reflector Sidelobe Degradation Due to Random Surface Errors," *IEEE Trans. Antennas Propagat.*, vol. AP-34 (February 1986): 164–172.
93. S. C. Wu and Y. Rahmat-Samii, "Average Pattern and Beam Efficiency Characterization of Reflector Antennas with Random Surface Errors," *JEWA*, vol. 5, no. 10 (1991): 1069–1087.
94. K. Bahadori and Y. Rahmat-Samii, "Characterization of Effects of Periodic and Aperiodic Surface Distortions on Membrane Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 9 (September 2005): 2782–2791.
95. R. P. Corkish, "A Survey of the Effects of Reflector Surface Distortions on Sidelobe Levels," *IEEE Antennas Propagat. Mag.*, vol. 32, no. 6 (December 1990): 6–11.
96. C. S. Powell, "Trends in Astronomy: Mirroring the Cosmos," *Sci. Am.*, vol. 265, no. 5 (November 1991): 112–123.
97. Y. Rahmat-Samii, "Array Feeds for Reflector Surface Distortion Compensation: Concepts and Implementation," *IEEE Antennas Propagat.*, vol. AP-32, no. 4 (August 1990): 20–26.
98. A. C. Schell, "The Multiple Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-14, no. 5 (September 1966): 550–560.
99. A. W. Rudge and N. A. Adatia, "Offset-Parabolic-Reflector Antennas: A Review," *Proc. IEEE*, vol. 66, no. 12 (December 1978): 1592–1618.
100. S.-W. Lee and Y. Rahmat-Samii, "Simple Formulae for Designing on Offset Multibeam Parabolic Reflector," *IEEE Trans. Antennas Propagat.*, vol. AP-29, no. 3 (May 1981): 472–478; Corrections: vol. AP-30, no. 2 (March 1982): 323.
101. R. J. Pogorzelski, "On the Numerical Analysis of the Subreflector of an Offset Cassegrain Microwave Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-32, no. 6 (June 1984): 595–601.

102. D.-C. Cheng and W. V. T. Rusch, "An Offset-Fed Reflector Antenna with an Axially Symmetric Main Reflector," *IEEE Trans. Antennas Propagat.*, vol. AP-32, no. 11 (November 1984): 1230–1236.
103. W. V. T. Rusch et al, "Derivation and Application of the Equivalent Paraboloid for Classical Offset Cassegrain and Gregorian Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-38, no. 8 (August 1990): 1141–1149.
104. Y. Rahmat-Samii, "Extension of Subreflector Increases Antenna Efficiency," *NASA Tech. Brief.*, vol. 11, no. 10, item 134; and JPL Invention Report NPO-16872/6391.
105. V. Krichevsky and D. F. DiFonzo, "Optimum Feed Locus for Beam Scanning in Offset Cassegrain Antennas," *IEEE Int. Symp. Antennas, Propagat.*, Quebec, Canada, *Digest*, vol. 11 (June 1980): 553–556.
106. V. Krichevsky and D. F. DiFonzo, "Optimum Feed Locus for Beam Scanning in the Symmetry Plane of Offset Cassegrain Antennas: Two-Dimensional Case," *COMSAT Tech. Rev.*, vol. 11, no. 1 (Spring 1981): 131–157.
107. V. Krichevsky and D. F. DiFonzo, "Beam Scanning in the Offset Gregorian Antenna," *COMSAT Tech. Rev.*, vol. 12, no. 2 (Fall 1982): 251–269.
108. V. Galindo, "Design of Dual-Reflector Antennas with Arbitrary Phase and Amplitude Distributions," *IEEE Trans. Antennas Propagat.*, vol. AP-12, no. 4 (July 1964): 175–180.
109. W. P. Williams, "High Efficiency Antenna Reflector," *Microwave J.*, vol. 8, no. 7 (July 1965): 79–82.
110. K. Miya (ed.), *Satellite Communications Technology*, Chap. 5, Sec. 5.4 (Tokyo: KDD Engineering and Consulting, Inc., 1980): 177–180.
111. G. L. James, "An Analysis of Radiation Pattern and G/T[A] for Shaped Dual Reflector Antennas," *IEEE Proc.*, vol. 127, pt. H, no. 1 (February 1980): 52–60.
112. W. V. T. Rusch, "Phase Error and Associated Cross-Polarization Effects in Cassegrainian-Fed Microwave Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-14, no. 3 (May 1966): 266–275.
113. J. J. Lee, L. I. Parad, and R. S. Chu, "A Shaped Offset-Fed Dual Reflector Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-27, no. 2 (March 1979): 165–171.
114. R. Mittra and V. Galindo-Israel, "Shaped Dual Reflector Syntheses," *IEEE Antennas Propagat. Soc. Newsletter* (August 1980): 5–9.
115. B. S. Westcott, F. A. Stevens, and F. Brickell, "Exact Synthesis of Offset Dual Reflectors," *Elect. Lett.*, vol. 16, no. 5 (February 28, 1980): 168–169.
116. V. Galindo-Israel et al, "On the Theory of the Synthesis of Offset Dual-Shaped Reflectors-Case Examples," *IEEE Trans. Antennas Propagat.*, vol. 39, no. 5 (May 1991): 620–626.
117. F. Watanabe and Y. Mizuguchi, "An Offset Spherical Tri-Reflector Antenna," *Trans. IECE Japan*, vol. E66, no. 2 (February 1983): 108–115.
118. S. Nomoto and F. Watanabe, "Shaped Reflector Design for Small-Size Offset Dual Reflector Antennas," *Electron. Commun. Japan*, vol. 72, Pt. 1, no. 11 (November 1989): 11–18.
119. J. Bergmann, R. C. Brown, P. J. B. Clarricoats, and H. Zhou, "Synthesis of Shaped-Beam Reflector Antenna Patterns," *IEEE Proc. H*, vol. 135, no. 1 (February 1988): 48–53.
120. B. Schlobohm and F. Arndt, "Small Earth Station Antenna Synthesized by a Direct PO Method," *Space Commun.*, vol. 7 (1990): 621–628.
121. Y. Rahmat-Samii and J. Mumford, "Reflector Diffraction Synthesis Using Global Coefficients Optimization Techniques," *IEEE AP-S Intl. Symp.* (1989): 1166–1169.
122. D. W. Duan and Y. Rahmat-Samii, "Reflector Antennas with Superquadric Aperture Boundaries," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 8 (August 1993): 1164–1167.
123. A. Michaeli, "Elimination of Infinities in Equivalent Edge Currents, Part I: Fringe Current Components," *IEEE Trans. Antennas Propagat.*, vol. AP-34, no. 7 (July 1986): 912–918.
124. P. Y. Ufimtsev, "Elementary Edge Waves and the Physical Theory of Diffraction," *Electromagn.*, vol. 11, no. 2 (April–June 1991): 125–160.
125. D. W. Duan and Y. Rahmat-Samii, "A Generalized Diffraction Synthesis Technique for High Performance Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 1 (January 1995): 27–40.

Chapter 16

Wideband Microstrip Antennas

L. Shafai

University of Manitoba

CONTENTS

16.1	INTRODUCTION	16-2
16.2	SUBSTRATE EFFECT ON BANDWIDTH	16-4
16.3	EFFECT OF PATCH SHAPE ON BANDWIDTH	16-5
16.4	PROXIMITY COUPLED MICROSTRIP ANTENNAS	16-6
16.5	APERTURE COUPLED MICROSTRIP ANTENNAS	16-7
16.6	STACKED MICROSTRIP PATCH ANTENNAS	16-15
16.7	MICROSTRIP ANTENNAS WITH L-SHAPED PROBE	16-24
16.8	SLOTTED WIDEBAND PATCH ANTENNAS	16-27
16.9	VARIATIONS OF U-SLOT PATCH ANTENNA	16-33
16.10	EBG LOADED SLOTTED PATCH ANTENNAS	16-42
16.11	MICROSTRIP PATCH ANTENNAS WITH EBG SUBSTRATE	16-45

16.1 INTRODUCTION

Microstrip antennas were born of microstrip circuit technology and inherited many characteristics, such as low radiating efficiency and narrow bandwidth, that are undesirable for a radiator. However, they offered many desirable features in terms of size, low profile, ease of integration with circuits, and forming arrays. Consequently, research has been conducted to improve their performance as radiators. Fundamentally, there are two types of microstrip antennas: the *traveling wave* type and the *resonant* type. The traveling wave microstrip antennas are for the most part some forms of printed lines that have been modified to radiate more efficiently. Examples are the comb lines, rampart lines, and other microstrip discontinuity lines.¹ These antennas generally have wide impedance bandwidths, but are negatively affected by the frequency dependence of their radiation characteristics such as the beam squint and sidelobe. The resonant microstrip antennas have relatively well-defined radiation characteristics, but suffer from narrow bandwidths. This class of microstrip antennas has received most of the attention in designs and bandwidth improvement. This chapter will provide a summary of the broadbanding techniques and examples of different designs capable of providing wideband characteristics. Since resonant microstrip antennas generally have broader bandwidths for their radiation characteristics like gain, sidelobe, and polarization, the bandwidth in this chapter will refer primarily to the impedance bandwidth of the antenna unless otherwise stated.

Resonant microstrip antennas usually take the form of a conducting printed patch. The patch shape can be arbitrarily determined, but the most common types are rectangular, square, circular, triangular, elliptical, rings, and their derivatives. Because a printed patch is usually over a conducting ground plane, it stores electric charge and energy and acts as a reactive circuit element. Consequently, the impedance bandwidth of a microstrip patch is typically very small, and depending on its substrate parameters and thickness, it can be less than 1 percent, or at best, a few percent. This range of bandwidths is smaller than the bandwidths of other antenna types and is insufficient for most current applications in communication and remote sensing, especially in new areas like microwave and medical imaging. Thus a need has arisen for broadening the bandwidth of microstrip patch antennas. Fortunately, researchers always respond to a challenge, and a number of effective techniques have been developed to increase the bandwidth of microstrip antennas from a few percent to around 50 percent and more, well beyond that of other antenna types. Consequently, microstrip patch antennas are no longer narrowband antennas. However, a price has been paid for broadening the bandwidth of the patch antenna, either in gain, polarization, or in the complexity of the antenna configuration. The drawbacks will be reviewed and discussed in relation with each wideband antenna type.

A number of recent books^{2-4,8} and articles⁵⁻⁷ have surveyed various broadbanding techniques for microstrip patch antennas. These have been referred to here for brevity in lieu of the original papers. There have been two main techniques for broadening the bandwidth of microstrip antennas: the *structural* technique and the *circuit theory* approach. The former deals mostly with the properties of the substrate over which the patch is etched or with the patch shape. The latter uses impedance matching techniques originally developed in circuit theory and then adapted to microwave networks, and here to antennas. However, the relationships are seldom obvious; the application of impedance matching techniques has not been systematic. This is mostly because the majority of broadbanding techniques were found and developed initially by antenna engineers using experimental procedures, and recently by use of commercially available software packages. This is unfortunate, because understanding the principles of impedance matching techniques in circuits can provide analytic methods to supplement the numerical method for determining the most suitable, or at least practically desirable, solutions.

The structural technique has been evolving. Initially, it involved primarily the antenna substrate. Recently, however, it has been expanded to include both the antenna substrate and its geometry, which have commonality with the circuit theory techniques. Microstrip patches are modified in shape and incorporated with slots of various sizes and shapes. Optimization codes have also been used with full wave analysis software to determine patch shapes for given applications. In reality, however, these techniques cause loading of the patch with impedance elements and are implicit applications of the impedance matching techniques.

The major innovation in the substrate area has come from applying the *electromagnetic band gap* (EBG) materials and metamaterials to microstrip antennas. In conventional substrates the ground plane electrical conductivity causes the image effect of the patch and limits its bandwidth and radiation efficiency. With these new materials, which are basically simulated materials using ordinary substrates, other electromagnetic boundary conditions, like a perfect magnetic conductor and a reactive surface, can be simulated. The image effect and the electrical properties of the patch are modified, and consequently the bandwidth and radiation efficiency of the antenna change. Naturally, these phenomena can also be explained and improved by the circuit theory approaches. However, they are easier to implement and understand by antenna engineers using electromagnetic theorems and boundary conditions.

In the sections to follow, antenna bandwidth is defined and related to its input parameters. Then the microstrip antenna bandwidth dependence on its substrate parameters is described. The impedance matching techniques using the circuit theory approach are presented briefly next. Because this method is most effective in enhancing the patch antenna bandwidth, the cases that follow are based on its applications. Sample examples using EBG materials are also provided. References are provided to help in reviewing other designs and innovations not listed in this chapter. Recently, there have been efforts to design ultra wideband (UWB) antennas for new UWB system applications. However, these antennas are required to have dipole-like omnidirectional patterns, not directive like microstrip antennas. For this reason they have not been discussed here.

The impedance bandwidth of a microstrip antenna is determined by its impedance characteristics. However, this is not a useful method in practical application, as the circuit model of the antenna is approximate and changes during the design and experimental tuning. A more practical approach is to define it in terms of its input parameters. The reflection coefficient Γ , or S_{11} , is readily measured by a network analyzer and is commonly used to define the antenna bandwidth. It is also a meaningful parameter, as it signifies the reflected power not available to the antenna, i.e., the *return loss*. Naturally, one is expected to minimize Γ . However, the antenna input impedance is a complex quantity that is also frequency dependent. The situation is more problematic in microstrip antennas because of the uncertainty due to fabrication tolerances in the substrate parameters. For these reasons the return loss of microstrip antennas is normally assumed satisfactory if it is better than 10 dB, and the antenna bandwidth is defined as

$$BW = \frac{(f_H - f_L)_{\Gamma=-10\text{dB}}}{f_o} \quad (16-1)$$

where

$$\Gamma = \frac{Z_{in} - Z_o}{Z_{in} + Z_o} \quad (16-2)$$

and Z_{in} and Z_o are the antenna and feed impedances respectively. Alternatively, the bandwidth is defined in terms of the antenna input VSWR in the form of

$$BW = \frac{VSWR - 1}{Q\sqrt{VSWR}} \quad (16-3)$$

where Q is the antenna quality factor, normally defined in terms of the stored energy and power loss as

$$Q = \frac{\text{Energy stored}}{\text{Power lost}} \quad (16-4)$$

In terms of the return loss, the antenna VSWR is given as

$$VSWR = \frac{1 + |\Gamma|}{1 - |\Gamma|} \quad (16-5)$$

A VSWR of less than 2 refers to a return loss of about 9.54 dB. A lower VSWR of about 1.5 is often desirable and relates to a return loss of about 14 dB. Much lower VSWR values and consequently, much lower return losses, may also be desirable in some applications. However, this can realistically be achieved if the substrate parameters are well defined and the antenna fabrication tolerances are kept low. For these reasons, in this chapter a return loss of 10 dB will be assumed for the antenna bandwidth definition.

16.2 SUBSTRATE EFFECT ON BANDWIDTH

When a patch antenna is etched directly on the dielectric substrate, the permittivity ϵ_r and thickness h of the dielectric influence its bandwidth. For small thicknesses the effects can be estimated intuitively, where the patch antenna can be viewed as a lossy capacitor. Thus the energy stored in the patch antenna increases by increasing the substrate ϵ_r and reducing its thickness h . This increases the Q of the patch antenna and reduces its bandwidth. These effects are shown in Figure 16-1. For $\epsilon_r = 2.2$, as the thickness h increases from

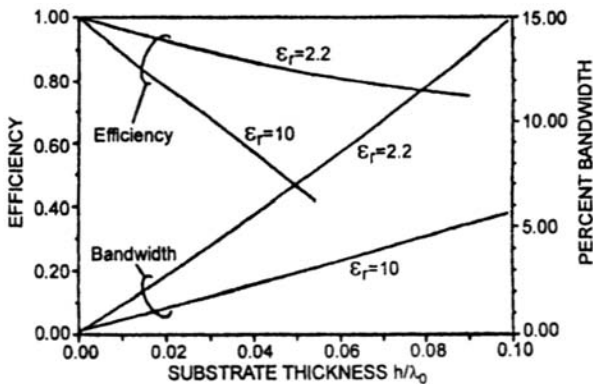


FIGURE 16-1 Effect of substrate thickness and dielectric constant on the impedance bandwidth (VSWR < 2) and radiation efficiency (after D. M. Pozar⁷ © IEEE 1992)

0 to $0.1 \lambda_0$, the bandwidth increases from 0 to 15 percent. For the same thickness, but $\epsilon_r = 10$, the rise in the bandwidth is only about 5.8 percent. The variation of the bandwidth is within expectation, but is not linear, due to other factors such as resistive and radiation losses.

In practice, decreasing the substrate permittivity is feasible and depends on the material selection criteria. However, increasing the thickness h is not as easy or desirable. Substrate modes become excited at discontinuities that scatter at other discontinuities radiating with undesirable polarization and radiation patterns. This phenomenon can become a source of serious performance degradation in microstrip arrays. Higher-order modes of the patch also become excited and contribute to the radiation, again degrading the antenna efficiency, polarization, and radiation patterns. In addition, increasing the substrate thickness also increases the antenna volume, thus increasing its weight and cost. Increasing the substrate thickness h , therefore, is not a good choice for enhancing the bandwidth.

The preceding substrate effects are valid for a patch over a ground plane, which is an electric conductor. The negative image of the patch currents on the ground plane degrades the antenna radiation efficiency and bandwidth. For instance, if the ground plane of the patch were replaced by a magnetic conductor generated using EBG materials, then the ground plane effects would reverse, and the patch antenna radiation efficiency and bandwidth would improve significantly. An example of such a design will be provided later in this chapter. The performance of a simple patch antenna designed to operate over a simulated magnetic ground plane will be compared with that of a conventional patch. Bandwidths in excess of 20 percent are obtained, which compares well with around 3 percent for the conventional patch antennas of similar design.

16.3 EFFECT OF PATCH SHAPE ON BANDWIDTH

The shape of the patch is its main parameter and naturally affects the antenna bandwidth. However, the possibilities are endless and the patch shape also affects its other electrical characteristics such as polarization and gain. It is therefore difficult to provide a general rule and guideline for the patch shape selection entirely based on its bandwidth. In addition, some of the patch shape deformations involve resonant slots that couple to the patch to increase the bandwidth. These configurations act as coupled resonators and are discussed later in this chapter. In this section we address the conventional patch shapes.

For solid patch shapes like square and circle, the bandwidths are nearly the same and proportional to their size. The deviations start when the shape changes significantly and becomes a narrow or wide rectangle. If the radiation edge becomes narrow, the radiation loss decreases and the antenna Q increases, reducing the bandwidth. The opposite is true for a patch with a wide radiating edge. More noticeable effects are found when the central part of the patch is removed to make a ring. Historically, annular ring (or square ring) antennas were considered to be wideband. This is true, however, only for the second mode of the patch, when it operates in the TM_{12} mode. For the first mode, the patch operates in the TM_{11} mode similar to a circular disk, and the antenna is a hybrid of a solid patch and a printed loop. For this mode the antenna bandwidth decreases rapidly as more central conductor is removed, making it a thinner ring.⁹

A more effective means of designing wideband microstrip antennas is by using coupled resonators, which can yield bandwidths in excess of 50 percent. The patch antenna need not be very special, but must be able to couple to another patch or to resonant elements. Resonant elements that are commonly used are a near resonant slot in the ground plane, a resonant slot on the patch itself, or a near resonant feed probe like a bent monopole. This technique stems from the lumped element circuit theory involving the impedance transformation,

specifically in coupled tuned circuits and transformers. If the resonant frequencies of two circuits are the same or nearly the same, then a critical coupling coefficient can be found to match two dissimilar impedances of the generator and load. Much larger bandwidths are obtained if the impedance matching is relaxed and other coupling coefficients are selected.

Note that in using the coupled resonator approach for broadbanding, it is possible to obtain an optimum configuration for wideband match but still fail to achieve the match. This situation usually arises from the presence of reactive components or from the dissimilarity of the patch and the feed impedances. The presence of a loop on the Smith Chart indicates coupled resonators, and the antennas have potential to be broadband. However, the loop may not be centered, due to an extra inductive or capacitive component. When the extra input reactance is removed, the loop moves to the chart center and improves the bandwidth.

16.4 PROXIMITY COUPLED MICROSTRIP ANTENNAS

As indicated earlier, increasing the height of the substrate increases the impedance bandwidth of the patch antenna. However, this causes additional problems, especially in using a transmission line feed. It increases the line width that together with the increased substrate height enhances the feedline radiation. This in turn causes difficulties in array design, where the feed network radiation deteriorates the antenna efficiency and polarization as its radiation is randomly polarized. A solution for this problem is found in using a proximity coupled feed, where the transmission line is etched on a separate lower substrate. In this manner, the parameters of the patch and its feed can be selected separately to optimize their individual performance. For instance, the patch radiation efficiency improves by lowering the permittivity and increasing the thickness of its substrate, but the opposite must be selected for the feedline.

The use of separate substrates for the patch and the feed also allows the use of matching stubs. Two different matching stubs are normally used, one in series with the line at its end, and the other orthogonal to the line before the patch. The end stub is in parallel with the patch input impedance and can influence its reactive component. Consequently, it can be used effectively to move the input impedance loop to the center of the Smith Chart, or to tune the resonance frequency of the patch. Also, its coupling to the resonant circuit of the patch provides a condition for broadbanding. This stub is also located below the patch and its radiation is negligible. The stubs orthogonal to the feedline can similarly be used for broadbanding.¹⁰ However, they are lossy due to the standing waves generated between the stub and patch. Also, the stub becomes a source of radiation that deteriorates the antenna polarization and efficiency. Therefore, it is preferred not to use them for improving the antenna impedance bandwidth.

The original configuration was investigated by Pozar and Kaufman¹¹ (see Figure 16-2). By using a circuit simulator and the experimental data, they showed that the microstrip line coupled to the patch is equivalent to a capacitor in series with the patch RLC equivalent circuit. The size of the impedance loop on the Smith Chart was controlled by the length of the series stub at the end of the line, and an orthogonal stub was used to move the loop to the chart center. Typical results are shown in Figure 16-3. When the tuning orthogonal stub is not connected, the input impedance at resonance is about $40\ \Omega$. Adding the tuning stub moves the impedance loop to the right, where it can be centered to maximize the bandwidth. In this example the height of both substrates was selected to be equal, $h = 0.158\text{ cm}$, and their permittivity was $\epsilon_r = 2.2$. The measured impedance plot gives a bandwidth of 13 percent for $\text{VSWR} < 2$ (3.375 GHz to 3.855 GHz). It was also shown that the orthogonal tuning stub strongly affects the cross polarization, and shorter stubs must be preferred.

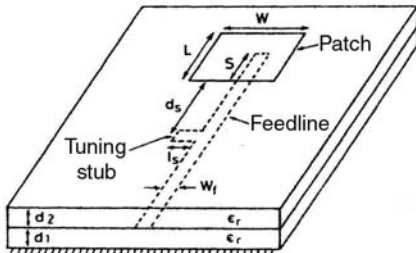


FIGURE 16-2 Geometry of proximity-coupled microstrip antenna: Substrate parameters: $\epsilon_r = 2.20$, $d_1 = d_2 = 0.158$ cm. Dimensions: $L = 2.5$ cm, $W = 4.0$ cm, $W_f = 0.5$ cm, $l_s = 0.65$ cm, $d_s = 3.3$ cm, $S = 1.25$ cm (after D. M. Pozar and B. Kaufman¹¹ © IEE 1987).

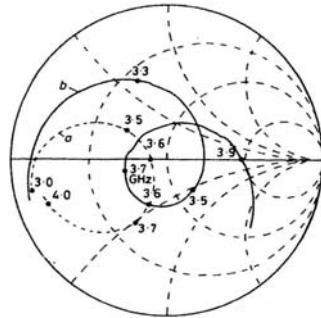


FIGURE 16-3 Smith Chart plot of measured impedance locus of the proximity-coupled microstrip antenna: (a) with a phase reference at edge of patch, without tuning stub and (b) with a phase reference at stub location, with tuning stub (matched antenna) (after D. M. Pozar and B. Kaufman¹¹ © IEE 1987)

In the preceding design a cross polarization of -20 dB was achieved. The performance of this antenna with different permittivities for the substrates and their heights was later studied in Splitt and Davidovitz.¹² Both rectangular and circular patches were considered, and their impedance bandwidths and efficiencies studied. It was shown that their impedance bandwidth can be increased to as high as 20 percent.

The performance of this microstrip antenna type was further investigated recently by shaping the end stub. In Tethakumary et al¹³ the end stub was bent orthogonally to form an L -shape with the feedline. Its impedance and radiation characteristics were studied experimentally, and an impedance bandwidth of around 20 percent (3.00–3.66 GHz) was achieved. The hook and T -shape^{14,15} stubs have also been investigated, but the bandwidth increase has been marginal. Wide bandwidths were also obtained without tuning stubs but by using parasitically coupled patches (see Legay and Shafai¹⁶). With two parasitic patches, bandwidths in excess of 15 percent at 4 GHz or high gains around 10.3 dBi were obtained and could be increased by adding additional patches. Figures 16-4 and 16-5 show the antenna geometry and its input impedance with two coupled patches.

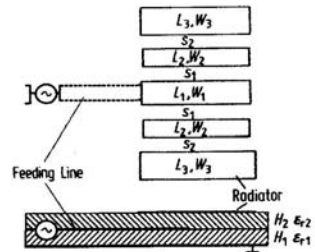


FIGURE 16-4 An EMC patch surrounded by parasitic elements (after H. Legay and L. Shafai¹⁶ © AEU 2002)

16.5 APERTURE COUPLED MICROSTRIP ANTENNAS

In the proximity coupled and single-layer transmission line feeds, both feedline and radiating patch are on the same side of the ground plane. These structures are topologically simple and are easier to fabricate or investigate. However, the feedline also radiates in the

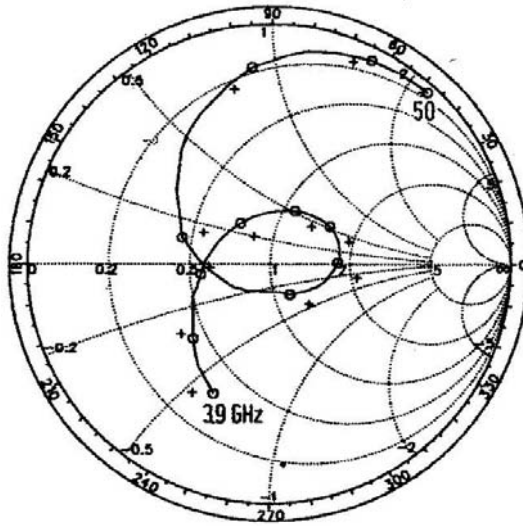


FIGURE 16-5 Input conductance of a wideband three coupled patch antenna: o-o-o Theory, + + + Experiment (after H. Legay and L. Shafai¹⁶ © AEU 2002). Parameters: $\epsilon_{r1} = \epsilon_{r2} = 2.5$, $H_1 = H_2 = 1.58$ mm, $L_1 = 22$ mm, $L_2 = 19$ mm, $W_1 = 8.4$ mm, $W_2 = 8.4$ mm, $s_1 = 2.1$ mm.

same direction as the patch and causes deterioration of the antenna gain and especially the polarization. The problem becomes more severe in large arrays, where the feed network grows complex and contains numerous power dividers and impedance matching sections. These junctions generate reflections, surface waves, and multiple resonant modes in the feed circuit. The surface waves, although evanescent, scatter from the nearby junctions and cause radiations with random phase and polarization. The reflected waves, resonant or otherwise, cause enhanced current distributions that increase the feed network loss and alter the power and phase distributions in the array. Consequently, the array gain and efficiency deteriorate and become frequency dependent. Similarly, its sidelobes and cross polarization also increase and become frequency dependent. Thus there is a need for a different feed type. The slot coupled feed is one such feed, where the entire feed network is moved below the ground plane and coupled to the radiating patch through a slot in the ground plane. However, the fabrication of the antenna becomes more complex and costly.

The original slot coupled feed was proposed and investigated by Pozar,¹⁷ where the coupling slot was selected to be small to prevent its radiation. The geometry of this type of antenna is shown in Figure 16-6. Since then, numerous investigators have studied the properties of this structure and have introduced variations in both shape and size of the coupling slot and the radiating patch to improve its impedance bandwidth, gain, and polarization. Zurcher¹⁸ investigated an inverted patch antenna. The patch was etched on a thin dielectric that acted as the protective cover when inverted. The combination was then separated from the ground plane by a thick foam substrate. The slot was enlarged to improve the bandwidth, which was fed by a microstrip line etched on a thin substrate below the ground plane. The antenna was called *SSFIP*, after its structure of Strip-Slot-Inverted Patch, as shown in Figure 16-7. An impedance bandwidth of 11.7 percent for the antenna was achieved that was increased to 21.1 percent for its 4×4 array.

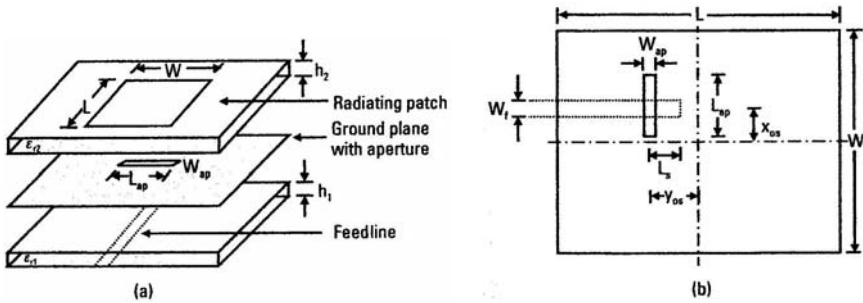


FIGURE 16-6 Aperture-coupled microstrip antenna: (a) exploded and (b) top views (after G. Kumar and K. P. Ray³ © Artech House 2003)

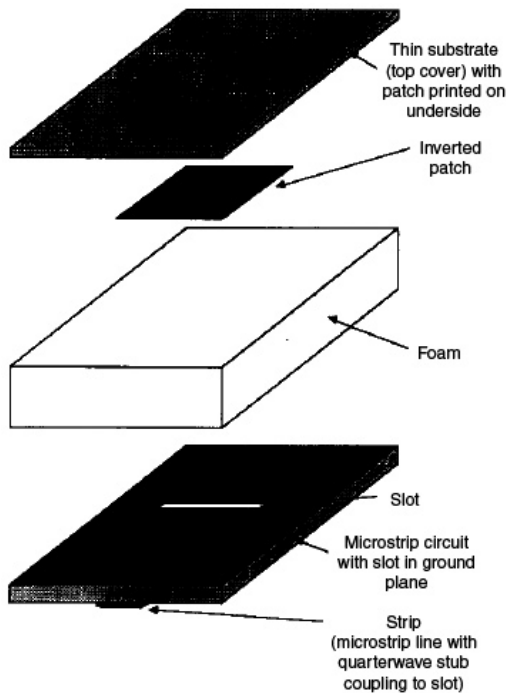


FIGURE 16-7 Exploded view of SSFIP antenna structure (after J. F. Zurcher¹⁸ © *Electronics Letters* 1988)

The coupling through the slot was shown to be dependent on the slot shape.^{19,20} In Pozar and Targonski¹⁹ both rectangular and dog-bone-shaped apertures were investigated. In the latter aperture, the rectangular slot was loaded at both ends by half-size transverse slots with rounded corners. The end loading of the slots extended the aperture electric field across the slot well into the slot ends. This extended strong electric field in the slot increases the coupling between the patch and the feedline. In other words, loading the slot ends

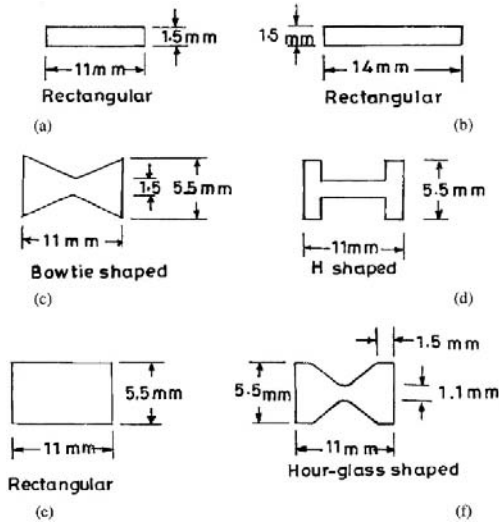


FIGURE 16-8 Shapes and dimensions of different apertures (after V. Rathi et al.²¹ © IEEE 1995)

increases its effective length. The coupling efficiencies of different apertures, namely narrow and wide rectangular apertures and H , bowtie, and hourglass shapes, were investigated in Rathi et al.²¹ They are shown in Figure 16-8, with dimensions in mm. The antenna was a rectangular patch with $L = 4$ cm, $W = 3$ cm, and the top and bottom substrates were the same with parameters as $\epsilon_r = 4.3$ and $h = 0.16$ cm. The width and length of the microstrip feedline were $W_f = 0.3$ cm and $L_s = 2$ cm. Measured input impedances showed that the end loading and end broadening increase the coupling. They had the opposite effect at the slot center. The bowtie performed better than the rectangular apertures, followed by the H and hourglass apertures. In the last case, the hourglass aperture combines the end loading of the H aperture with the end broadening of the bowtie, thus providing the strongest coupling.

Further bandwidth enhancement was obtained by using nearly resonant apertures.²² In this manner, the problem is essentially that of two coupled resonators: the coupling aperture and patch. The input impedance plot of the antenna, with its parameters at 5.6 GHz, is shown in Figure 16-9. For $VSWR < 1.5$, the achieved bandwidth is about 23 percent, which becomes nearly 29 percent for $VSWR < 2$. This is considerably higher than the bandwidths achieved by previous configurations and also maintains the quality of its radiation patterns, shown in Figure 16-10 for the copolar component, and in Figure 16-11 for the cross-polar component. Also, within the band the measured gain remained at or above 8 dBi. The significant performance of this design is the fact that its measured copolar contours are symmetric. The measured cross-polarization contours are also symmetric and very low and about -40 dB at the boresight; they remain very low in a large angular range of about 25° about the main beam axis. The cross polarizations in the diagonal planes are maximum at about -15 dB but at large angles off the main beam. These characteristics are due to a strong excitation of the microstrip patch dominant TM_{11} mode and to weak excitation of the higher-order modes.

A further bandwidth enhancement was obtained by using a CPW feedline²³ (see Figure 16-12). It is interesting to note that the coupling aperture is placed coplanar, on the CPW line, and the CPW external ground planes act as the reflecting ground of the patch antenna.

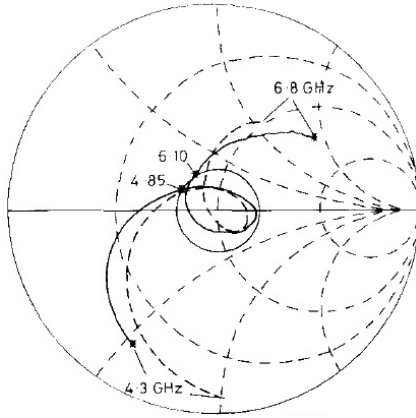


FIGURE 16-9 C band resonant aperture coupled microstrip antenna impedance locus (after F. Croq and A Papiernik²² © *Electronics Letters* 1990):
 ----- Theoretical. ——— Measured. Feed: $\epsilon_{rf} = 2.2$, $\tan \delta = 0.001$, $H_f = 0.762$ mm; $W_f = 2.32$ mm, $L_s = 2.85$ mm. Slot: $A_w = 0.8$ mm, $A_l = 15.4$ mm. Square patch: $W_p = 17$ mm, $H_p = 5.5$ mm, $\epsilon_{rp} = 1$. Radome: $H_s = 1.6$ mm, $\epsilon_{rs} = 2.2$, $\tan \delta = 0.001$.

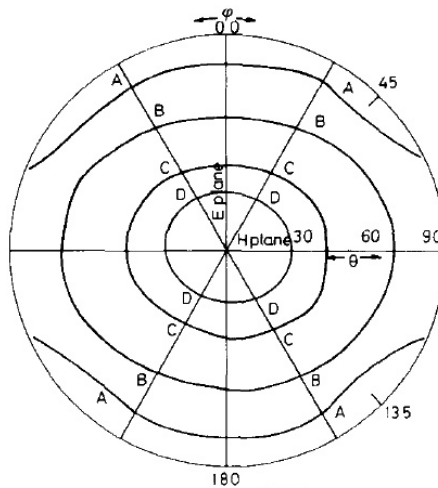


FIGURE 16-10 Copolar contour plots at 5.6 GHz of the resonant aperture coupled microstrip antenna (after F. Croq and A Papiernik²² © *Electronics Letters* 1990): $\theta = 0^\circ - 90^\circ$; $D = -1.0$ dB; $C = -3.0$ dB; $B = -10$ dB; $A = -20$ dB. The antenna dimensions are mentioned in Figure 16-9.

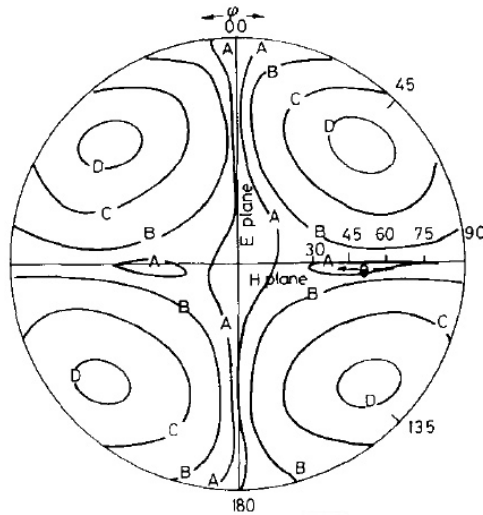


FIGURE 16-11 Cross-polar contour plots at 5.6 GHz of the resonant aperture coupled microstrip antenna (after F. Croq and A. Papiernik²² © *Electronics Letters* 1990): $\theta = 0^\circ$ – 90° ; $D = -40$ dB; $C = -30$ dB; $B = -20$ dB; $A = -15$ dB. The antenna dimensions are mentioned in Figure 16-9.

Consequently, this antenna needs only one substrate, and its configuration is simpler. The tuning stub is also placed coplanar with the line and at its end. The CPW structure also makes the active element integration much simpler. A moment method optimization of the antenna provided a wide impedance bandwidth of about 36 percent, at 3.4 GHz, as shown in Figure 16-13. This is larger than the bandwidth achieved by a microstrip line feed, discussed earlier. However, its cross polarization is larger. The measured radiation patterns are shown in Figure 16-14, showing cross polarization around -25 dB, and a front-to-back ratio of 12 dB. The computed gain was 9 dBi.

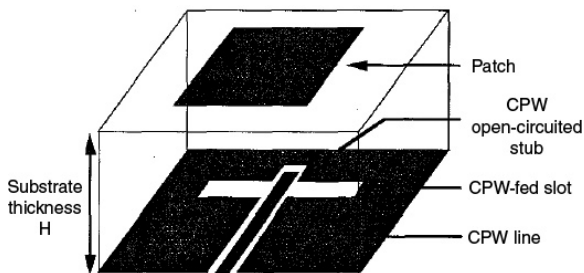


FIGURE 16-12 CPW-fed aperture coupled microstrip antenna (CPWFA) tuned to open-circuited stub (after Giauffret et al²³ © *Electronics Letters* 1995)

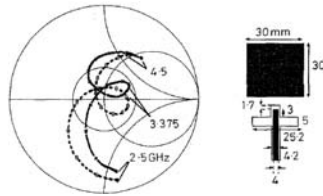


FIGURE 16-13 Measured and theoretical input impedance of the CPWFA in Figure 16-12 (after Giauffret et al²³ © *Electronics Letters* 1995): $H = 9.5$ mm, $\epsilon_r = 1.04$, $\tan \delta = 0.001$. -x-x- Experiment, -o-o- Theory.

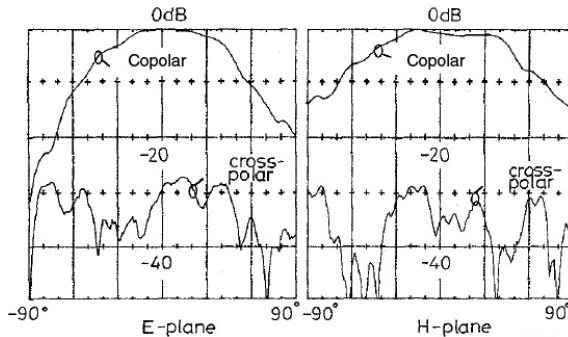


FIGURE 16-14 Measured copolar and cross-polar patterns of CPWFA in Figure 16-12 (after Giauffret et al²³ © *Electronics Letters* 1995)

Designs with ring and U -slot apertures have also been attempted, showing improved impedance bandwidths. An example is shown in Figure 16-15 and Figure 16-16, which uses a ring slot for aperture coupling.²⁴ The patch and slot configurations are complex in shape and use notches for optimization. The feedline end is wider and forms a patch-shaped stub. The configuration was optimized using the moment method software Ansoft Ensemble v8.²⁵ The simulated and measured VSWR results are compared in Figure 16-17, which also shows the antenna dimensional parameters. The H -shaped patch was etched on a thin 1.6-mm FR4 substrate, with a relative permittivity of 4.3. A 50- Ω microstrip line with a tuning stub is located on the lower surface of the lower substrate. A 10-mm air-filled substrate separates the ground plane and H -shaped patch. The structure consists of two resonators: the H -shaped patch and the ring slot. A measured bandwidth of 50.6 percent (2.46 – 4.09 GHz) for VSWR < 2 was achieved. The measured gain remained above 6 dBi in the entire band, as shown in Figure 16-18, and the cross polarization in both principal planes was below -25 dB. However, the measured back radiation was large and about -10 dB below the main beam. To reduce the back radiation, a second ground plane was placed behind the first, with an air

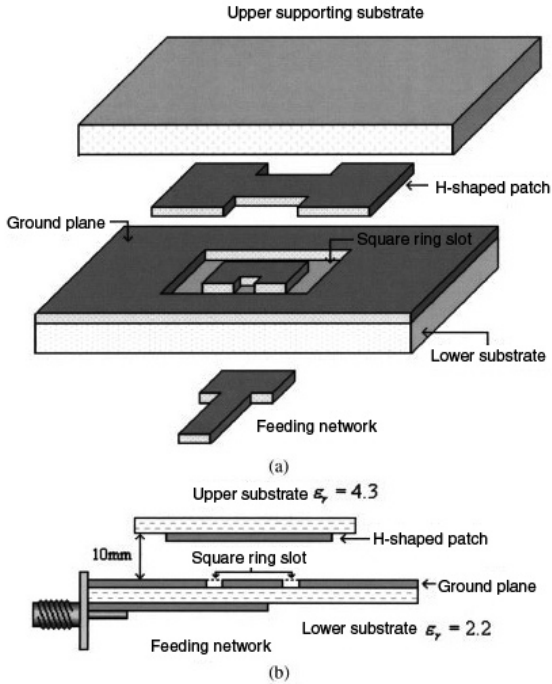


FIGURE 16-15 (a) Geometry of the antenna with a ring slot aperture; (b) Side view of the square-ring-slot coupled patch antenna (after Y. Qin et al²⁴ © *Microwave and Optical Technology Letters* 2005)

gap of 5.6 mm. With this configuration the antenna impedance bandwidth deteriorated to 44 percent. This indicates that the back radiation is the reason for enhanced bandwidth, which may also be the case for other aperture coupled antennas with large coupling apertures.

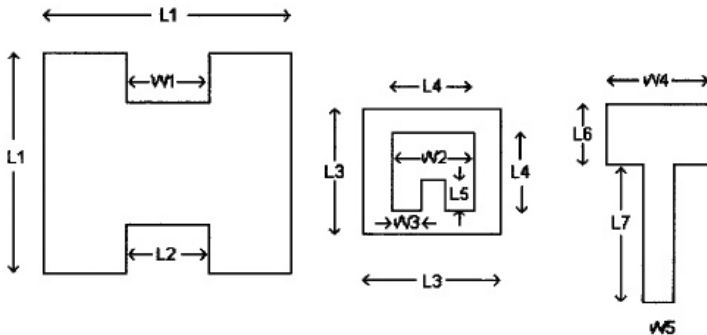


FIGURE 16-16 Detailed dimensions of the antenna in Figure 16-15 (after Y. Qin et al²⁴ © *Microwave and Optical Technology Letters* 2005)

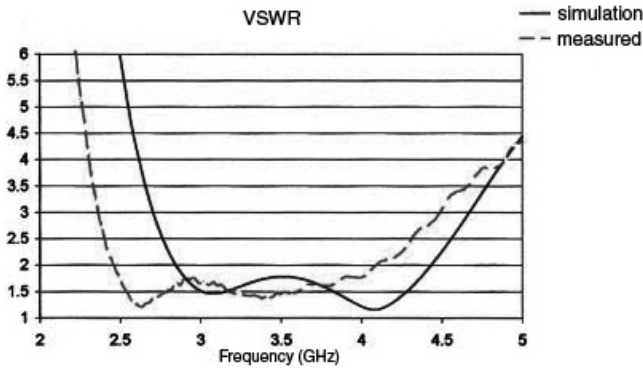


FIGURE 16-17 Predicted and measured VSWR of the antenna in Figure 16-15 (after Y. Qin et al²⁴ © *Microwave and Optical Technology Letters* 2005)

Dimensions of the design (in mm):

$L1 = 28$	$L2 = 9.5$	$L3 = 23$	$L4 = 14.8$
$L5 = 6.5$	$L6 = 10.6$	$L7 = 15.7$	$W1 = 9.5$
$W2 = 6.5$	$W3 = 5.7$	$W4 = 11.8$	$W5 = 3.012$

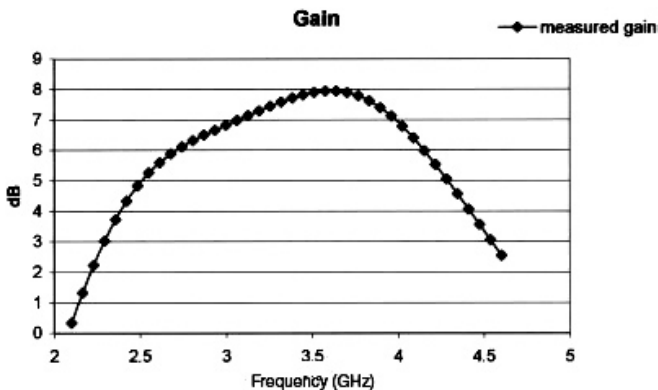


FIGURE 16-18 Measured gain vs. frequency of the antenna in Figure 16-15 (after Y. Qin et al²⁴ © *Microwave and Optical Technology Letters* 2005)

16.6 STACKED MICROSTRIP PATCH ANTENNAS

Among various wideband microstrip patch antennas, the coupled patch antennas are simpler to understand and design. Each patch is a resonator and equivalent to an RLC network. Thus two adjacent patches can be viewed as two coupled resonators and can be designed for wideband operation by using the circuit theory approach available in literature and discussed briefly at the beginning of this chapter. For instance, if the resonant frequencies of two coupled patches are nearly the same, then a wide impedance bandwidth can be obtained simply by adjusting their coupling to find the optimum coupling level, which is a single parameter optimization. You should recognize, however, that such a wideband design will not be the true optimum since the parameters of each patch, i.e., each resonator, may not be optimum. In practice, you must optimize the entire antenna geometry, including

the parameters of the coupled patches, together with their coupling level. Nevertheless, the technique of one parameter optimization is a very useful approach in designing coupled patch antennas. When only two stacked patches are used, then a simple adjustment of their vertical spacing provides a quick method of broadbanding their input impedance. In such cases only the lower patch is fed, and the size of the upper patch is made slightly larger to increase its coupling to the lower one.

Coupled patch antennas can be made of two or multiple patches. The case of two stacked coupled patches can be viewed as a single wideband microstrip patch antenna that can be used alone or in an array. Its gain will be slightly larger than that of a single-layered patch, due to its increased height above the ground plane, and controlled by its value in wavelength; that is, the larger the height of the stacked configuration, the larger its gain. This is the secondary benefit from the stacked structure, in addition to its broadband impedance bandwidth. The case of multiple coupled patch antennas can increase significantly both the input impedance bandwidth and the gain.²⁶ However, since such structures are coupled multiple resonators, they can have multiple resonances, some of which can be undesirable. For instance, different patches can have induced surface currents in opposite directions, reducing the gain and increasing the sidelobes. Since the coupling is frequency dependent, such phenomena are also frequency dependent and cause rapid variation of the antenna gain and radiation patterns. Thus multiple coupled patch antennas are significantly more difficult to design and must be investigated carefully.

The coupled patch antennas have fundamentally two different configurations, the planar structures and stacked configurations. Both these configurations were investigated concurrently.^{27,28} In the planar configuration the coupling is from the edges of the adjacent patches. Since the geometry represents two coupled resonators, the coupling coefficient is the most important parameter for broadbanding of the input impedance. However, the coupling in this case is normally weak and below the coupling level necessary for optimum broadbanding. Thus the degree of bandwidth broadening is limited to around 10 percent, unless multiple planar patches are used. Nevertheless, this type of impedance bandwidth increase becomes attractive in the design of planar arrays, like microstrip Yagi arrays, where the planar structure makes this type of impedance broadening very natural. In other cases this configuration becomes appealing only if there is a limitation on the antenna height and adequate planar surface is available for the coupled patches.

The stacked coupled patch structures have been favored in most applications because of the ease with which the coupling coefficient can be controlled by adjusting their separation height. Their other parameters such as the size of the patches and substrate parameters can also be modified independently to facilitate the designs. Consequently, much research has been done on these antennas, and impedance bandwidths of 10–30 percent for $VSWR < 2$ have been achieved. From broadband network theory, if the coupling between the two patches is too strong, i.e., above the required critical coupling, there are two resonance frequencies that increasingly separate by increasing the coupling coefficient. Since in stack patches their spacing controls the coupling, the antenna becomes dual band when patches are too close to each other. As the spacing is increased, the coupling reduces and the two resonances merge at the critical coupling to provide a broadband antenna. Further increase of the spacing reduces the bandwidth gradually.

In practice, the feed network and the substrate parameters also influence the antenna circuit parameters and the bandwidth. Normally, the feed, such as a coaxial probe or an aperture in the ground plane, is also a resonant circuit. Below resonance, they are equivalent to a reactive element and do not introduce additional resonances; however, they affect the circuit parameters of the stacked patches or move the input impedance curve of the antenna on the Smith Chart away from its center. A typical geometry of a coaxial probe-fed stacked microstrip patch antenna is shown in Figure 16-19a,²⁹ where the probe feed inductance is in series with the equivalent circuit of the antenna. If additional matching stubs should not

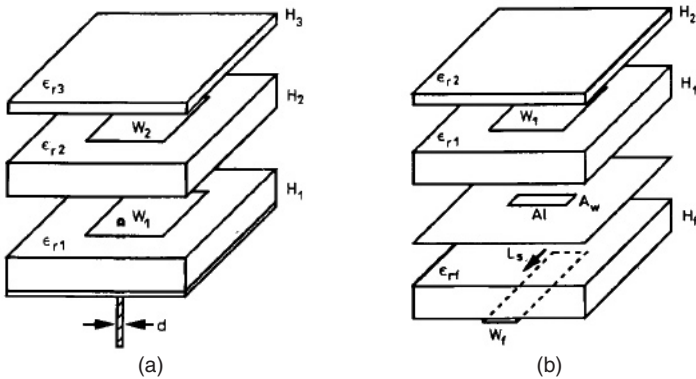


FIGURE 16-19 Exploded views: (a) dual patch coaxially fed antenna and (b) aperture coupled (after F. Croq and A. Papiernik²² © *Electronics Letters* 1990)

be used to cancel this reactance, then the antenna parameters must be changed to achieve a broadband. Alternatively, the feed location with respect to the patch center may be changed to partially compensate for the reactance. If, on the other hand, the feed size is large and near, or at resonance, within the antenna operating band, then the coupling to the antenna can further increase the input impedance bandwidth. This case corresponds to the near resonance aperture sizes in the aperture coupled microstrip antennas, discussed previously. With conventional probes this phenomenon does not occur, due to small substrate heights of conventional microstrip antennas. It can be introduced by bending the probe to reduce its vertical height. This configuration is also capable of broadband operation and will be discussed in a later section.

From earlier discussions, the lower patch requires stronger coupling and must be on a thin dielectric substrate (thickness $\leq 0.04\lambda_0$), which will also reduce the problems associated with thick substrates, like surface wave losses, weight, and cost. The upper patch requires weaker coupling and should be on a thicker low permittivity substrate ($\epsilon_r \approx 1$ and thickness $\approx 0.06\lambda_0$),³⁰ and is normally suspended in the air or placed on a foam substrate. In such cases, and since optimum couplings require a reasonable separation between the two patches, the upper patch is effectively over a substrate with a permittivity of close to unity. The design process, therefore, becomes somewhat easier. A good starting point is to design the lower patch using the low frequency limit, or for a slightly higher frequency, using conventional design equations. The upper patch in most cases is a few percent larger. The coupling between the two patches pushes the resonant frequency of the antenna toward the required limits. Such a design will be a good starting point for numerical optimization by use of full wave analysis codes. Figure 16-20 shows the effect of the lower and upper patch sizes on the resonance loop on the Smith Chart. Their effects are opposite. As the lower patch size increases, the loop moves to the left toward lower impedance, but for the upper patch it moves to the right and higher impedance.³⁰ Bandwidths in excess of 25 percent are obtained. A larger bandwidth of 40 percent is obtained using a CPW feed of Figure 16-21,³¹ which is nearly the same as the CPW-fed aperture coupled antenna of Figure 16-12.

A drawback of the stacked microstrip patch antennas is the direct radiation from the probe, which becomes significant at high frequencies at large lower substrate heights, especially in cross polarization. This phenomenon can be better understood by comparing the results for both probe and aperture feeding types. Figure 16-19 shows the geometry of these antennas,

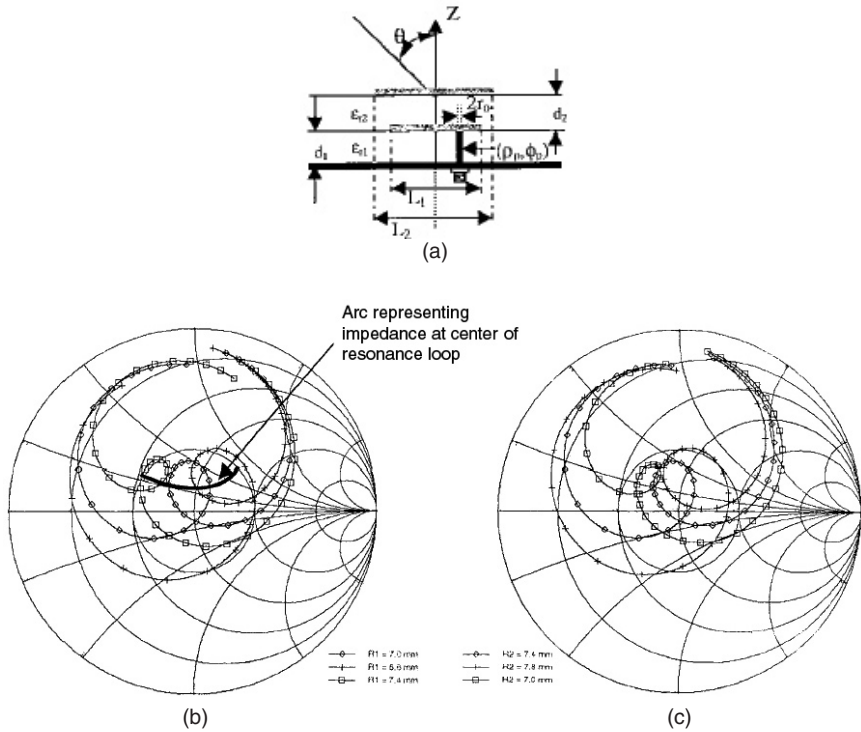


FIGURE 16-20 (a) Cross-sectional views of probe-fed rectangular stacked-patch geometry: Effects of patch dimensions on impedance locus (frequency range: 6–11 GHz). Parameters: $L_1 = 13.5$ mm, $W_1 = 12.5$ mm, $\epsilon_{r1} = 2.2$, $d_1 = 1.524$ mm, $\tan \delta_1 = 0.001$, $L_2 = 15$ mm, $W_2 = 16$ mm, $\epsilon_{r2} = 1.07$, $d_2 = 2.5$ mm, $\tan \delta_2 = 0.001$, $x_p = 5.4$ mm, $r_o = 0.325$ mm. (b) R_1 variation ($R_2 = 7.4$ mm). (c) R_2 variation ($R_1 = 7.0$ mm) (after R. B. Waterhouse³⁰ © IEEE 1999).

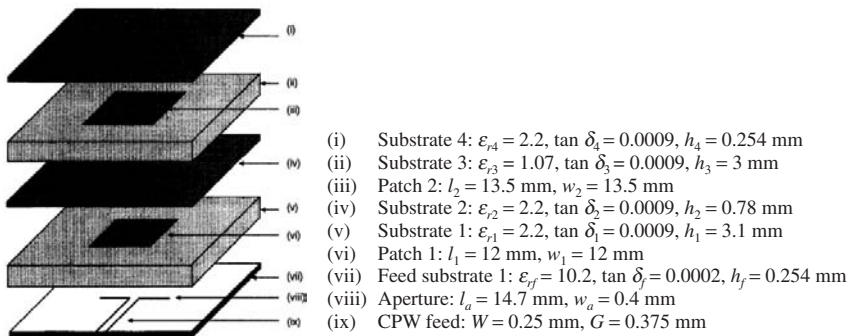


FIGURE 16-21 Geometry of broadband CPW-fed patch antenna (after W. S. T. Rowe and R. B. Waterhouse³¹ © Electronics Letters 1999)

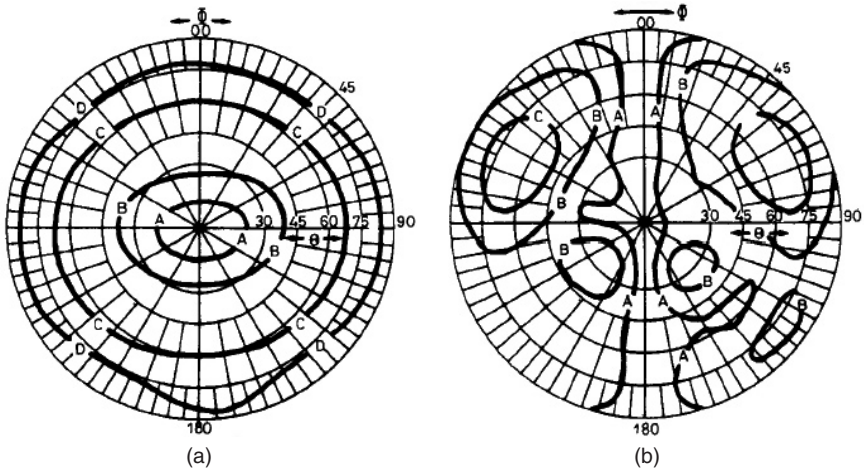


FIGURE 16-22 Coaxially fed dual patch antenna in Figure 16-19a: X-band antenna, radiation pattern at 12 GHz (measured): (a) Copolar component: $A = -1$ dB, $B = -3$ dB, $C = -5$ dB, $D = -10$ dB. (b) Cross-polar component: $A = -30$ dB, $B = -20$ dB, $C = -12$ dB. Antenna parameters: $W_1 = 8$ mm, $W_2 = 6.9$ mm; probe position: $X_p = Y_p = 2.5$ mm, $\epsilon_{r1} = 2.2$, $\epsilon_{r2} = 1$, $\epsilon_{r3} = 2.2$, $H_1 = 1.53$ mm, $H_2 = 1.53$ mm, $H_3 = 3.09$ mm (after F. Croq et al²⁹ © IEE 1993).

which are a probe-fed stacked patch antenna and a single-layer aperture coupled patch antenna.²⁹ They have impedance bandwidths in excess of 30 percent at 12 GHz. Figure 16-22 and Figure 16-23 provide the measured radiation contours for the copolar and cross-polar components of these feeds at the X-band. The probe-fed antenna shows asymmetry in the contours of both copolar and cross-polar components, due to the probe radiation. They are symmetric for the aperture-coupled antenna.

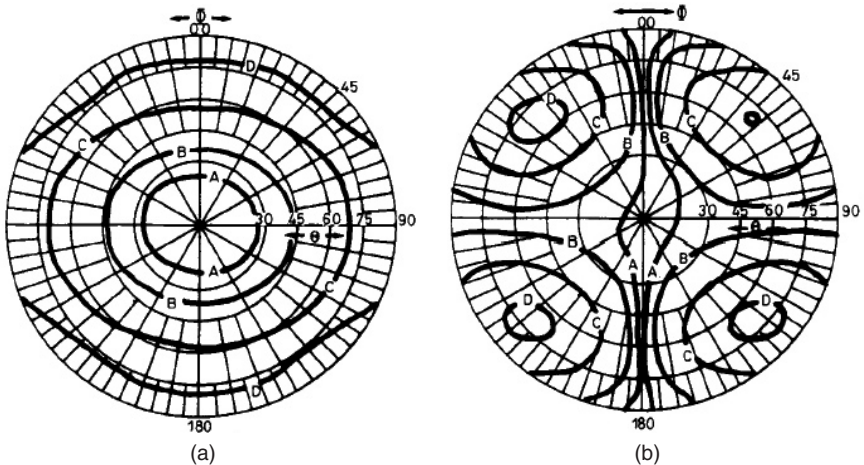


FIGURE 16-23 Aperture fed (one patch) antenna in Figure 16-19 and (b) X-band antenna, radiation pattern at 12 GHz (measured): (a) Copolar component: $A = -1$ dB, $B = -3$ dB, $C = -1$ dB, $D = -20$ dB. (b) Cross-polar component: $A = -35$ dB, $B = -28$ dB, $C = -22$ dB, $D = -17$ dB. Antenna parameters: $W_1 = 7$ mm, $A_1 = 8$ mm, $A_w = 0.4$ mm, $W_f = 4.9$ mm, $L_s = 2.4$ mm, $\epsilon_{rf} = 2.2$, $\epsilon_{r1} = 1$, $\epsilon_{r2} = 2.2$, $H_f = 1.54$ mm, $H_1 = 2.5$ mm, $H_2 = 1.54$ mm (after F. Croq et al²⁹ © IEE 1993).

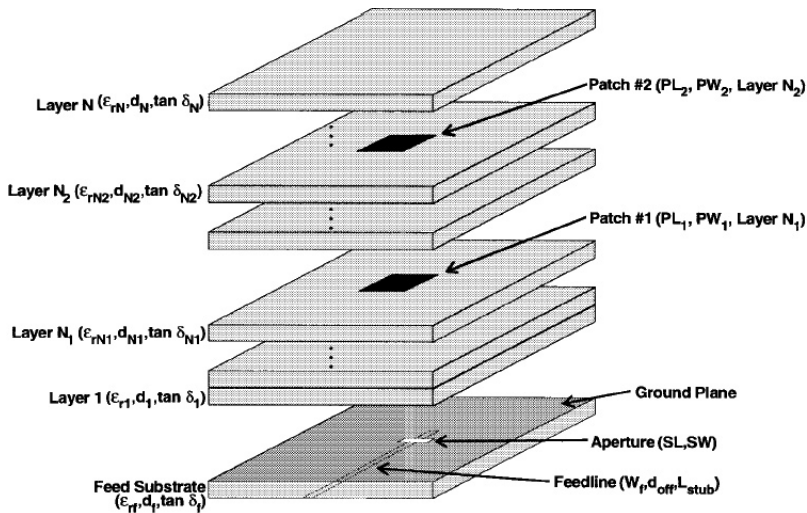


FIGURE 16-24 Geometry of multilayered aperture coupled stacked patch antenna (after S. D. Targonski et al.³² © IEEE 1998)

For further enhancement of the bandwidth, additional resonances are needed that can be introduced by adding more patches. In the stacked configuration the antenna height further increases. A more suitable configuration is to combine the aperture coupled geometry with stacked patches. Such a configuration is shown in Figure 16-24 and investigated in Targonski et al.³² The electromagnetic couplings are between three resonators consisting of a near resonant aperture and two stacked patches. In this design the aperture is coupled to the lower patch, and its size must be selected appropriately to increase the bandwidth through the correct value of the coupling coefficient and the resonance frequencies. The individual coupling between the aperture and the lower patch and that between the patches were investigated by a parametric study of their dimensions. The results are shown in Figure 16-25 to 16-27. The best coupling levels are shown in Figures 16-25*b*, 16-26*b*, and 16-27*b*, along with their dimensional parameters.

Similar studies were also conducted for the substrate thicknesses of the first and second layers and are shown in Figure 16-28 and Figure 16-29. These results show the interrelationship between the coupling of the resonators and the dimensional parameters that can be used to arrive at a desirable design solution, discussed in detail in Targonski et al.³² An example design was then selected to maximize the bandwidth. Its computed and measured return loss and gain are shown in Figures 16-30 and 16-31, which also show the antenna design data. The computed and measured bandwidths were 67 percent (5.05–10.1 GHz) and 69 percent (5.07–10.38 GHz). Considering the complexity of the antenna geometry, the agreement between the computed and measured results is satisfactory. Over most of the band, the gain is above 7 dBi and drops rapidly near the band edges. The computed front-to-back ratio ranged from 8 to 14 dB over most of the band.

Multiple resonances can also be established in dual-layer patch antennas by using multiple planar patches on the second layer.²⁶ The geometry of such an antenna is shown in Figure 16-32, where a single rectangular patch is placed on the lower substrate. On the upper substrate four rectangular patches are etched, such that when symmetrically located on the lower patch, they overlap at the corners.

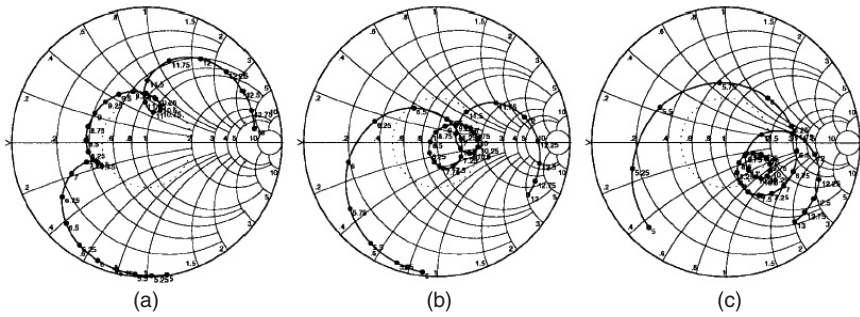


FIGURE 16-25 Input impedance as a function of aperture length: (a) $SL = 8$ mm, (b) $SL = 10$ mm, and (c) $SL = 12$ mm. Other parameters: $W_f = 4.75$ mm, $L_{\text{stub}} = 3.4$ mm, $SW = 0.8$ mm. Patches: $PL_1 = PW_1 = 9.1$ mm, $N_1 = 1$, $PL_2 = PW_2 = 9.1$ mm, $N_2 = 2$. Dielectric layers: $N = 3$, $\epsilon_{r1} = 2.33$, $\tan \delta_1 = 0.0012$, $d_1 = 1.6$ mm, $\epsilon_{r1} = 2.2$, $\tan \delta_1 = 0.0009$, $d_1 = 3.175$ mm, $\epsilon_{r2} = 1.07$, $\tan \delta_2 = 0.0009$, $d_3 = 3$ mm, $\epsilon_{r3} = 2.2$, $\tan \delta_3 = 0.0009$, $d_3 = 0.127$ mm. (after S. D. Targonski et al.³² © IEEE 1998).

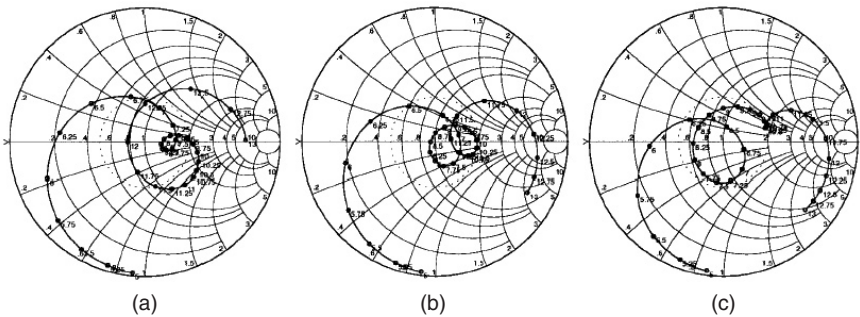


FIGURE 16-26 Input impedance as a function of lower patch size: (a) $PL_1 = PW_1 = 8$ mm, (b) $PL_1 = PW_1 = 9$ mm, and (c) $PL_1 = PW_1 = 10$ mm (after S. D. Targonski et al.³² © IEEE 1998)

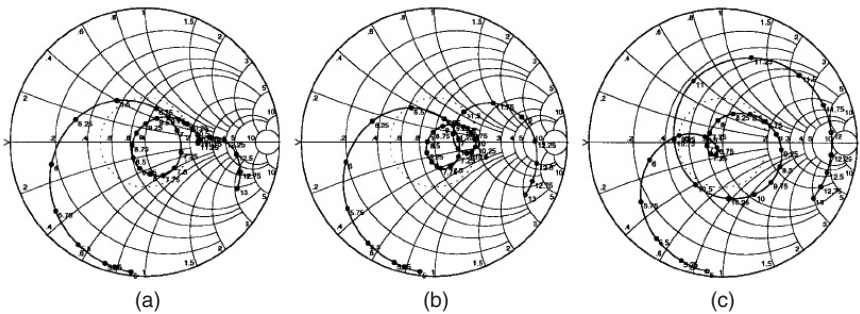


FIGURE 16-27 Input impedance as a function of upper patch size: (a) $PL_2 = PW_2 = 9$ mm, (b) $PL_2 = PW_2 = 10$ mm, and (c) $PL_2 = PW_2 = 12$ mm (after S. D. Targonski et al.³² © IEEE 1998)

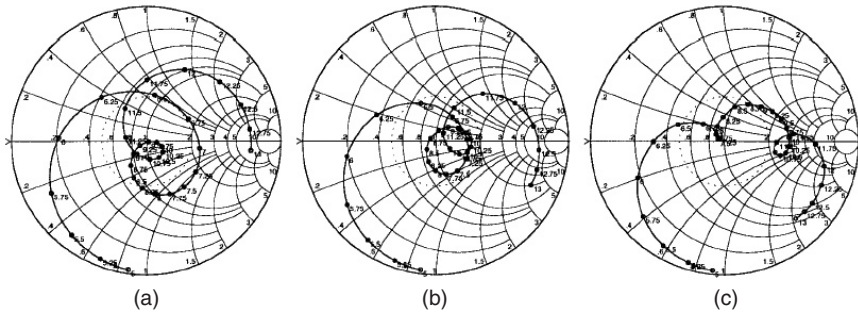


FIGURE 16-28 Input impedance as a function of Layer # 1 dielectric thickness: (a) $d_1 = 2.5$ mm, (b) $d_1 = 3$ mm, and (c) $d_1 = 4$ mm (after S. D. Targonski et al³² © IEEE 1998)

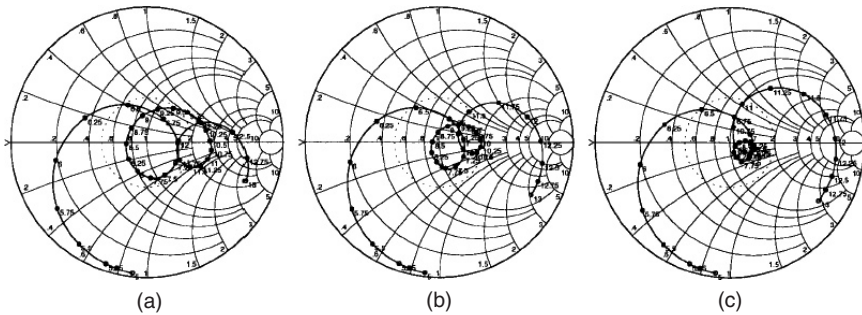


FIGURE 16-29 Input impedance as a function of Layer # 2 dielectric thickness: (a) $d_2 = 2$ mm, (b) $d_2 = 3$ mm, and (c) $d_2 = 4$ mm (after S. D. Targonski et al³² © IEEE 1998)

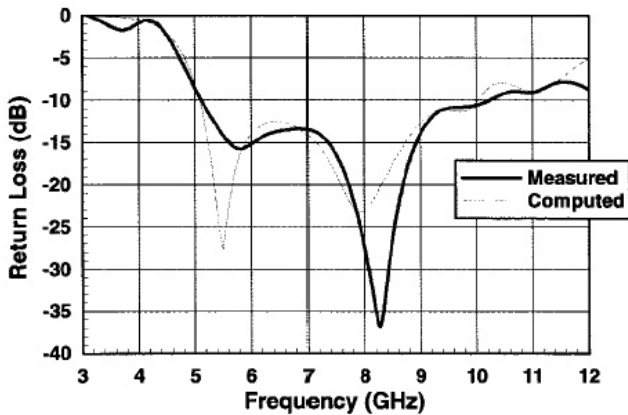


FIGURE 16-30 Computed and measured return loss for aperture coupled stacked patch antenna: Feed parameters: $W_f = 0.5$ mm, $d_{\text{off}} = 5.4$ mm, $L_{\text{stub}} = 5.8$ mm, $\epsilon_{rf} = 2.2$, $\tan \delta_f = 0.0009$. Aperture: $SL = 16$ mm, $SW = 1$ mm. Antenna substrate: $N = 4$, $\epsilon_{r1} = 1.07$, $d_1 = 1.2$ mm, $\tan \delta_1 = 0.0009$; $\epsilon_{r2} = 2.2$, $d_2 = 3.175$ mm, $\tan \delta_2 = 0.0009$, $\epsilon_{r3} = 1.07$, $d_3 = 3.1$ mm, $\tan \delta_3 = 0.0009$, $\epsilon_{r4} = 2.53$, $d_4 = 0.508$ mm, $\tan \delta_4 = 0.003$. Patch elements: $PL_1 = 10.8$ mm, $W_1 = 20$ mm, $N_1 = 2$, $PL_2 = 11.2$ mm, $PW_2 = 20$ mm, $N_2 = 3$ (after S. D. Targonski et al³² © IEEE 1998).

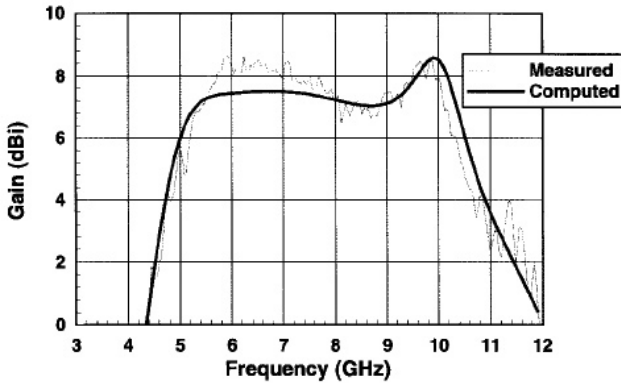


FIGURE 16-31 Computed and measured gain for aperture coupled stacked patch antenna. Antenna parameters are mentioned in Figure 16-30 (after S. D. Targonski et al.³² © IEEE 1998).

The degree of coupling between the lower patch and the upper patches is controlled by the amount of the overlap area, as well as by the spacing between the lower and the upper patches. The latter condition is similar to dual stacked patches, where a low permittivity foam substrate is used to separate the two patches. The level of coupling between the lower and one of the upper patches is small, but the combined coupling levels become sufficient to reach the critical coupling coefficient needed for the broadband input impedance. For the configuration of

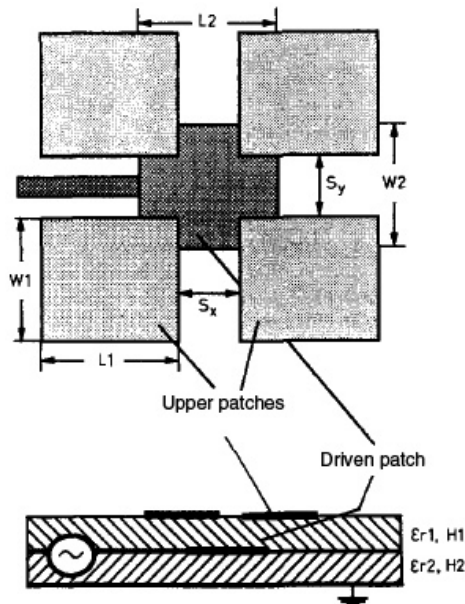


FIGURE 16-32 Geometry of the stacked microstrip antenna consisting of electromagnetically coupled four patches to a driven patch (after H. Legay and L. Shafai²⁶ © IEE 1994)

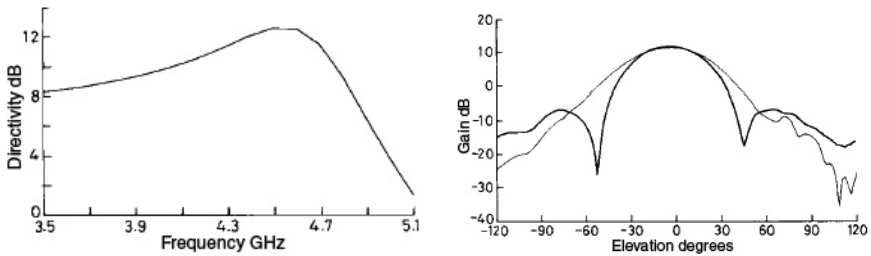


FIGURE 16-33 Radiation patterns for the antenna in Figure 16-32 at 4.3 GHz: (a) Directivity. (b) Measurement. Substrate parameters: $\epsilon_{r1} = \epsilon_{r1} = 2.55$, $h_1 = h_2 = 1.58$ mm. Upper patches: $L_1 = 20$ mm, $W_1 = 16$ mm. $S_X = S_Y = 16$ mm. Lower patch: $L_2 = 21.3$ mm, $W_2 = 20$ mm (after H. Legay and L. Shafai²⁶ © IEE 1994).

Figure 16-32 the computed and measured impedance bandwidths were 20 percent. One added benefit of this configuration is its high gain of about 12 dBi. In fact, the original antenna in Legay and Shafai²⁶ was designed to provide high gain without an array. In microstrip arrays the feed network losses deteriorate its performance in sidelobes, cross polarization, gain, and the noise temperature. In the configuration of Figure 16-32, the entire feed network is eliminated to remove its detrimental effect. This stacked configuration was designed as a high gain element for a thinned active microstrip array. Its high gain and wide bandwidth allowed considerable simplification in the feed network of a large array, with improved efficiency and electrical performance. Its measured gain is shown in Figure 16-33.

Further increase in the bandwidth can be obtained by adjusting the spacing between the lower patch and upper patches. In this case, the level of coupling depends on the corner overlaps as well as the spacing. In the bandwidth optimization the two parameters must be adjusted simultaneously to maintain the needed coupling levels. Bandwidths in excess of 45 percent have been obtained. The antenna gain was found to be high for large overlaps and lower interlayer spacing. It is decreased by reducing the overlap or increasing the second substrate height. It was found that the radiation from the lower patch was out of phase with those of the upper patches, and that it tends to reduce their radiation intensity, and thus the antenna gain. Any means of reducing its radiation tends to improve the gain. Increasing the overlap or reducing the interlayer spacing results in covering the central patch and thus reducing its radiation.

16.7 MICROSTRIP ANTENNAS WITH L-SHAPED PROBE

Coaxial probe feeding is attractive in many practical applications, especially in single antennas, because of its simplicity and ease of fabrication. However, wide impedance bandwidths are obtained only by using stacked patches, and the probe height is usually too small to be resonant. Longer probes become too inductive and also radiate in the dipole mode, which is in cross polarization. Thus with longer probes the antenna must become capacitive to compensate for the probe inductance. These properties were generated by a bent *L*-shaped probe,^{33,34} shown in Figure 16-34. A long probe is used as a feed and is bent below the patch. The horizontal section of the probe couples capacitively to the patch and compensates for its inductance. In this manner a resonant probe length can be selected to couple to the patch and increase its bandwidth. With a rectangular patch a parametric study was carried out to investigate its bandwidth and gain. The results are shown in Figure 16-35. Bandwidths in the order of 36 percent, centered at 4.5 GHz, are obtained. The copolar gain is frequency dependent, but remains high over most of the band. It drops rapidly near edges of the band. For better antenna performance the vertical section of the probe must fall slightly outside the patch, i.e., $S < 0$. The radiation patterns are

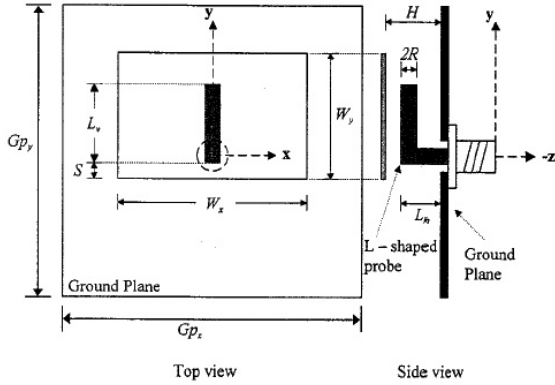
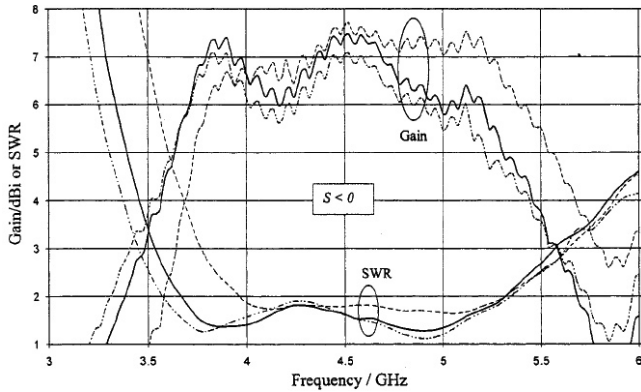
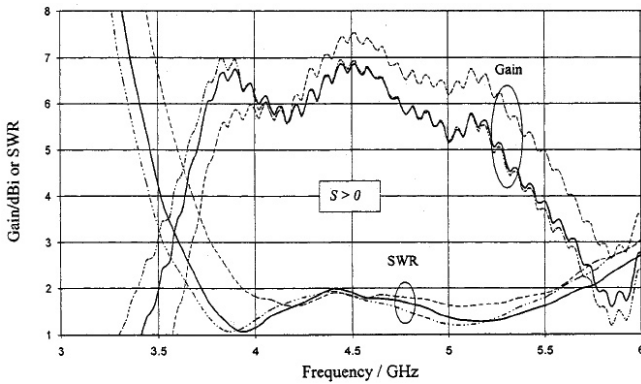


FIGURE 16-34 Basic geometry of the L -shaped probe-fed patch antenna and the coordinate system (after C. L. Mak et al¹³³ © IEEE 2000)



(a)



(b)

$H = 5 \text{ mm}$ -----, $H = 6.6 \text{ mm}$ ———, $H = 7 \text{ mm}$ -.-.-.-

FIGURE 16-35 SWR and gain curves for the six cases in Table 16-1 (after C. L. Mak et al¹³³ © IEEE 2000)

TABLE 16-1 Bandwidth and Gain of the *L*-Shaped Probe-Fed Patch Antenna with Different Values of Parameters

Set	H/mm	L_v/mm	L_H/mm	S/mm	Bandwidth (SWR ≤ 2)	Average Gain
1	5 ($0.075\lambda_o$)	10.42 ($0.1563\lambda_o$)	4.44 ($0.0666\lambda_o$)	-1.5 ($-0.0225\lambda_o$)	28%	7dBi
2	6.6 ($0.099\lambda_o$)	10.5 ($0.1575\lambda_o$)	5.68 ($0.0852\lambda_o$)	-3 ($-0.0450\lambda_o$)	36%	7dBi
3	7 ($0.105\lambda_o$)	9.88 ($0.1482\lambda_o$)	6.2 ($0.0930\lambda_o$)	-1 ($-0.0150\lambda_o$)	39%	6.5dBi
4	5 ($0.075\lambda_o$)	7.64 ($0.1146\lambda_o$)	4.8 ($0.0720\lambda_o$)	+1.5 ($0.0225\lambda_o$)	32%	6.5dBi
5	6.6 ($0.099\lambda_o$)	5.18 ($0.0777\lambda_o$)	6.48 ($0.0972\lambda_o$)	+2 ($0.0300\lambda_o$)	42%	6dBi
6	7 ($0.105\lambda_o$)	6.98 ($0.1047\lambda_o$)	6.68 ($0.1002\lambda_o$)	+1 ($0.0150\lambda_o$)	42%	6dBi

$W_x \times W_y = 30 \times 25 \text{ mm}^2, R = 0.5 \text{ mm}$

shown in Figure 16-36. The copolar patterns have nearly symmetric shape, but the cross polarization is asymmetric and increases near the band edges, causing the gain loss. In the region between 15° and 30° , off the main beam, it increases to around -15 dB .

The performance of the antenna with an *L*-shaped probe over cylindrical and conical grounds was also investigated.³⁵ The bandwidth improved to about 48 percent (2.01 to 3.2 GHz), and peak gain increased to 8.6 dBi. The cross polarization remained high at about -14.2 dB in the *H*-plane. The *L*-shaped probe was similarly investigated with circular and triangular patch

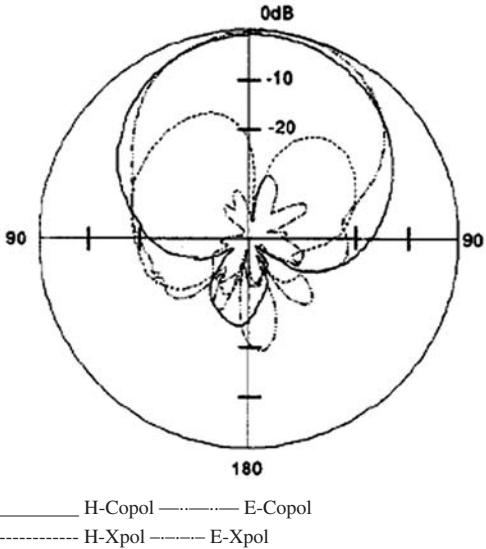


FIGURE 16-36 Radiation patterns of the antenna using parameters in set 2 of Table 16-1 at 4.53 GHz (after C. L. Mak et al³³ © IEEE 2000)

antennas,³⁶ and similar performances were obtained. In lieu of an L -shaped probe, the T -shaped probe has also been investigated.³⁷ Its performance in bandwidth, gain, and cross polarization was found to be similar to the L -shaped probe. However, adding a second T -shaped probe, symmetrically under the patch, reduced the cross polarization to below -20 dB.

16.8 SLOTTED WIDEBAND PATCH ANTENNAS

In aperture coupled patch antennas, the coupling between a near resonant slot and the patch broadened the bandwidth to about 30 percent or higher. Another location for the slot is the patch surface. However, since the slot is long, the patch must be wide and the slot must be bent but remain symmetrical on the patch. Using this concept a U -shaped slotted patch was experimentally investigated in Huynh and Lee,³⁸ and an impedance bandwidth of 47 percent, from 0.812 to 1.282 GHz, was obtained. Later, the same configuration was investigated numerically.³⁹ It was found that within the impedance band the antenna polarization changes at the upper frequencies. The realistic bandwidth, for the same polarization, was found to be about 30 percent. Even then the cross polarization was increased significantly near the end of its impedance band.^{39,40} Figure 16-37 shows the geometry of this antenna and its dimensional parameters. The slotted patch is over a thick foam substrate, and a coaxial probe is used to feed it at the patch center. Figure 16-38a shows its return loss plot, and Figure 16-38b the gain at two orthogonal polarizations. It is clear that the antenna polarization starts switching around 1 GHz.

Attempts were also made to design this antenna on a dielectric substrate $\epsilon_r > 1$. However, these initial attempts were not too successful, and impedance bandwidths were limited to around 20 percent. Instead, numerous authors have investigated successfully the same and derivatives of this antenna. In Bhalla and Shafai⁴¹ approximate expressions for minimum return loss frequencies of the U -slot antenna were proposed and investigated. They are given in Eqs. 16-6 and 16-7:

$$f_1 = \frac{c}{\left(\frac{W_s}{2} + W + \frac{3}{2} \Delta W - \frac{b}{2} \right) \sqrt{\epsilon_{eff}}} \quad (16-6)$$

$$f_1 = \frac{c}{\left(\frac{L}{2} + W + \frac{3}{2} \Delta W - \frac{W_s}{4} + F - a \frac{b}{2} + \frac{t}{2} \right) \sqrt{\epsilon_{eff}}} \quad (16-7)$$

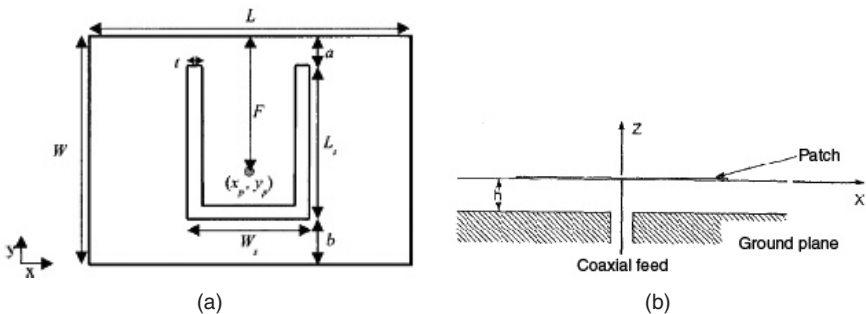
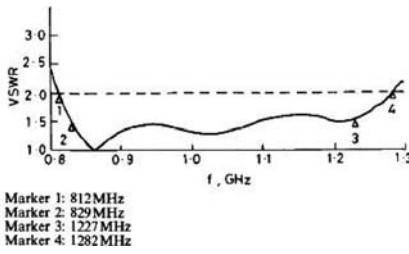
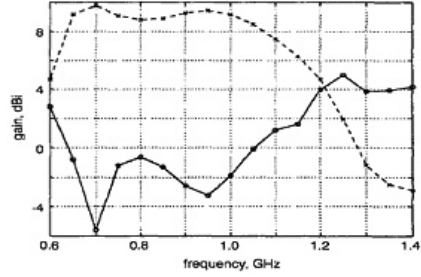


FIGURE 16-37 Geometry of coaxially fed rectangular patch with a U -shaped slot: (a) top view and (b) side view (after R. Bhalla and L. Shafai⁴¹ © *Microwave and Optical Technology Letters* 2002)



(a)



(b)

FIGURE 16-38 (a) VSWR against frequency for the coaxially fed rectangular patch with a U-shaped slot on foam. Dimensions of the antenna in centimeters: $W = 12.45$, $L = 21.97$, $W_s = 6.86$, $L_s = 8.22$, $a = 2.29$, $b = 1.94$, $t = 1.02$, $h = 2.69$, $F = W/2$ (after T. Huynh and K. F. Lee³⁸ © *Electronics Letters* 1995). (b) Maximum gain of x- and y-polarizations of the U-slot antenna.³⁹

where c is the velocity of light in free space and ΔW is the fringing effect calculated from

$$\Delta W = 0.412h \frac{(\epsilon_{eff} + 0.3) \left(\frac{L}{h} + 0.264 \right)}{(\epsilon_{eff} - 0.258) \left(\frac{L}{h} + 0.8 \right)} \quad (16-8)$$

and ϵ_e is the effective permittivity calculated from

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + \frac{10h}{W} \right)^{-\frac{1}{2}} \quad (16-9)$$

The parameters in these equations are defined in Figure 16-37 and Figures 16-39 through 16-41. The rationale for choosing these equations is based on observation of the

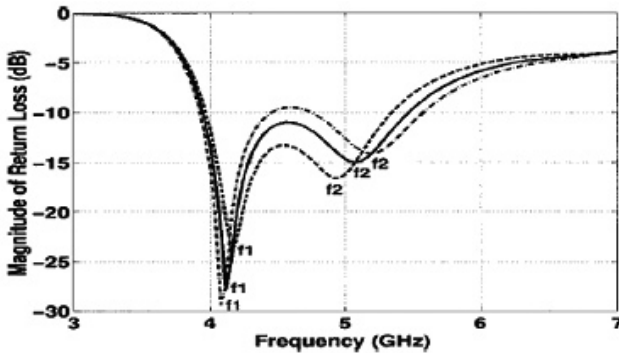


FIGURE 16-39 Return loss of the microstrip antenna with a single U-shaped slot:⁴¹ $L = 37.5$ mm, $W = 26$ mm, $L_s = 19.5$ mm, $W_s = 12$ mm, $a = 3.7$ mm, $b = 2.8$ mm, $t = 2.1$ mm, foam height $h = 5$ mm. ----- $F = 14.50$ mm, ——— $F = 15.0$ mm, -.-.- $F = 15.5$ mm.

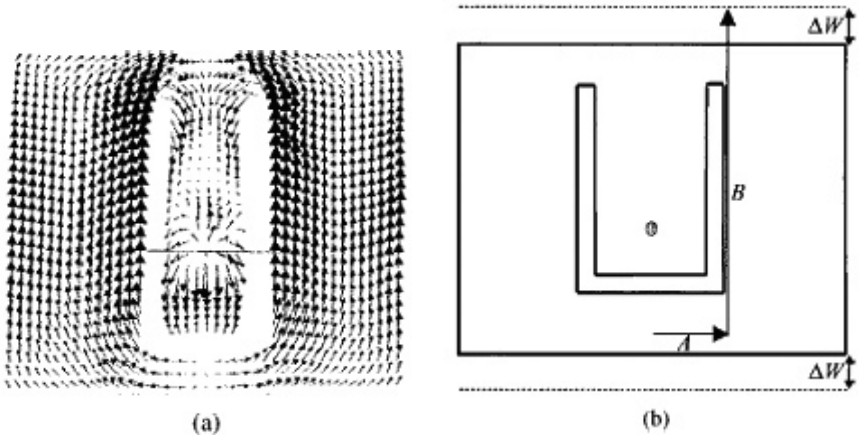


FIGURE 16-40 Surface-current distribution on microstrip patch at $f_1 = 4.12$ GHz (a) and (b) surface-current line on the patch at $f_1 = 4.12$ GHz. Here $F = 15.0$ mm. Other dimensions are the same as in Figure 16-39 (after R. Bhalla and L. Shafai⁴¹ © *Microwave and Optical Technology Letters* 2002).

patch current distributions at the null frequencies f_1 and f_2 , shown in Figure 16-40 and Figure 16-41. At f_1 , currents appear to start at the center point behind the slot and terminate at the opposite edge. These currents resemble the patch current distribution due to the TM_{11} mode. Thus, the length of the average current path will be half wavelength. At the second null the currents appear to start at the feed point and end at the center point behind the slot. This length is one wavelength at f_2 . These equations may be used for calculating approximately the null frequencies that can be used in determining the initial antenna dimensional parameters. However, since the substrate height is not present in these equations, they will only provide the patch and slot parameters.

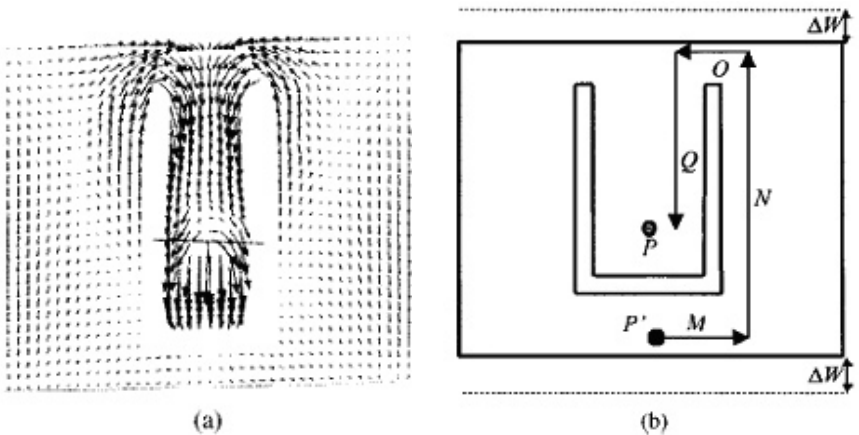


FIGURE 16-41 Surface-current distribution on microstrip patch at $f_2 = 5.08$ GHz (a) and (b) surface-current line on the patch at $f_2 = 5.08$ GHz. Here $F = 15.0$ mm. Other dimensions are the same as in Figure 16-39 (after R. Bhalla and L. Shafai⁴¹ © *Microwave and Optical Technology Letters* 2002).

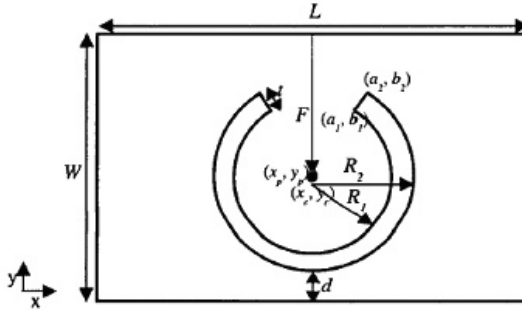


FIGURE 16-42 Geometry of the microstrip patch antenna with a circular arc slot (after R. Bhalla and L. Shafai⁴² © IEEE 2002)

A rectangular patch with a circular slot was also studied and designed.⁴² It provided a bandwidth of 31.1 percent, at a center frequency of about 4 GHz. Its geometry with dimensional parameters is shown in Figure 16-42. The return loss plot is shown in Figure 16-43. Again, approximate expressions for the return loss null frequencies were developed and are given next.

$$f_1 = \frac{c}{2 \left(R_1 + (W + 2\Delta W) - \frac{(\Delta W + d)}{2} \right) \sqrt{\epsilon_{eff}}} \quad (16-10)$$

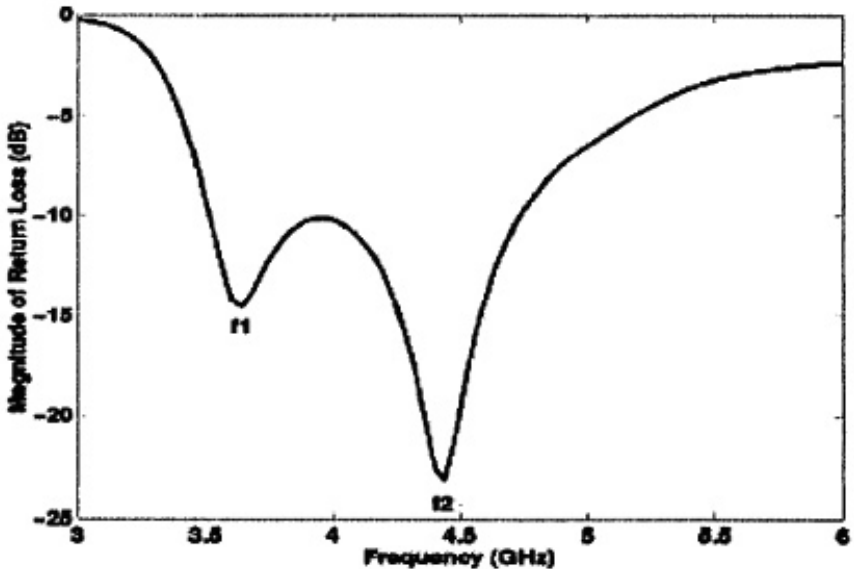


FIGURE 16-43 Return loss of the microstrip patch antenna with a circular arc slot, shown in Figure 16-42: $L = 55.5$ mm, $W = 29$ mm, $(x_c, y_c) = (0$ mm, -2.5 mm), $(a_1, b_1) = (5$ mm, 2.5 mm), $(a_2, b_2) = (7$ mm, 5.5 mm), $t = 2.83$ mm, $d = 1.37$ mm, $F = 16$ mm, foam substrate height = 5.5 mm (after R. Bhalla and L. Shafai⁴² © IEEE 2002).

$$f_2 = \frac{c}{\left(\frac{L}{2} + \frac{W}{2} + \frac{3}{2}\Delta W + \frac{R_2}{2} + F - \frac{d}{2}\right)\sqrt{\epsilon_{eff}}} \quad (16-11)$$

These equations were also developed using the current distributions at the null frequencies and found to be accurate within 4 percent.

A more elaborate, but approximate, design approach for a U -slotted rectangular patch is developed in Weigand et al⁴³ based on a detailed numerical investigation. They have defined four resonant frequencies: f_{res1} , f_{res2} , f_{res3} , and f_{res4} . They are related to the slot resonance, the TM_{01} mode of the patch, the antenna center frequency, and the slot resonance frequency in free space. The first resonance frequency is not radiating efficiently and is discarded. The second, third, and fourth resonant frequencies are related to the antenna dimensional parameters and the substrate relative permittivity. Based on these, a design approach is developed that, although approximate, helps in determining the initial design of the antenna. A final tuning may be necessary to yield the desired antenna. An important feature of this study is the fact that it includes the substrate permittivity in the design, and a wideband antenna on a dielectric substrate with $\epsilon_r = 2.2$ is designed that provides a bandwidth of 39 percent, between 4.21 and 6.26 GHz. Figure 16-44 defines the antenna dimensional parameters, and the approximate equations for the last three resonance frequencies are given by

$$f_{res2} \approx \frac{c}{2\sqrt{\epsilon_{eff}}\left(B + 2\Delta B + \frac{D}{2} - E\right)} \quad (16-12)$$

$$f_{res3} \approx \frac{c}{2\sqrt{\epsilon_{eff}}(B + 2\Delta B)} \quad (16-13)$$

$$f_{res4} \approx \frac{c}{(2C + D) + \sqrt{\epsilon_{eff(pp)}}(B - E - H + 2\Delta_{B-E-H})} \quad (16-14)$$

where ΔB and Δ_{B-E-H} are the fringing effect at the patch edge and the slot, respectively. Based on these equations, a design procedure is developed, which is described next.

1. Specify the center frequency and 2:1 VSWR bandwidth of the desired antenna. Approximate the center frequency as f_{res3} and the lower and upper frequency bounds of the bandwidth as f_{res2} and f_{res4} , respectively.
2. Select a substrate permittivity ϵ_r and a substrate thickness T . There is a lower limit on T below which broadband operation is unlikely. Therefore, the substrate thickness and permittivity should satisfy the following rule of thumb derived from the existing literature and parametric studies:

$$T \geq 0.06 \frac{\lambda_{res3}(\text{air})}{\sqrt{\epsilon_r}}$$

3. Estimate the quantity $B + 2\Delta B$ as follows:

$$B + 2\Delta B \approx \frac{c}{2\sqrt{\epsilon_r}f_{res3}}$$

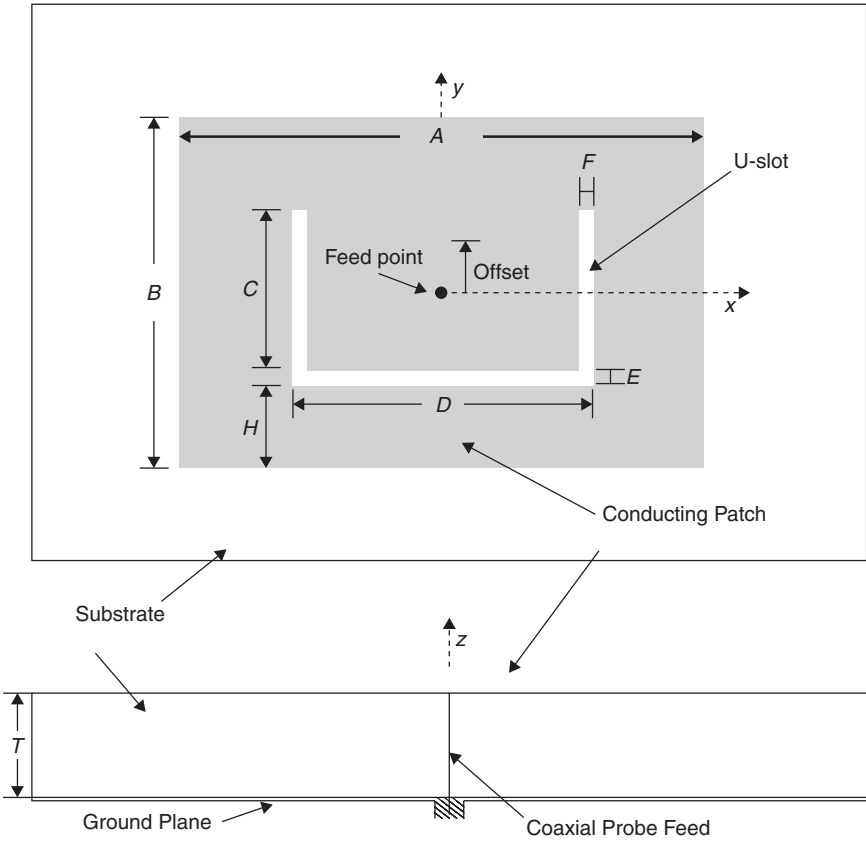


FIGURE 16-44 Geometry of the rectangular *U*-slot microstrip patch antenna (after S. Weigand et al⁴³ © IEEE 2003): Dimensions and material properties of the initial *U*-slot antenna used in the parametric studies (after S. Weigand et al⁴³ © IEEE 2003):

<i>A</i> (mm)	<i>B</i> (mm)	<i>C</i> (mm)	<i>D</i> (mm)	<i>E</i> (mm)	<i>F</i> (mm)	<i>H</i> (mm)	<i>R</i> (mm)	<i>T</i> (mm)	Offset (mm)	ϵ_r
36.0	26.0	12.0	16.0	2.0	2.0	4.0	0.5	5.0	0.0	2.20

4. Calculate *A* as

$$A = 1.5(B + 2\Delta B)$$

5. Calculate ϵ_{eff} and $2\Delta B$ using the following common equations:

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + \frac{12T}{A} \right)^{-1/2}$$

$$2\Delta B = 0.824T \frac{(\epsilon_{eff} + 0.3) \left(\frac{A}{T} + 0.262 \right)}{(\epsilon_{eff} - 0.258) \left(\frac{A}{T} + 0.813 \right)}$$

6. Back calculate the value of B :

$$B = \frac{c}{2\sqrt{\epsilon_{eff}} f_{res3}} - 2\Delta B$$

7. Select a starting value of slot thickness using the following rule of thumb:

$$E = F = \frac{\lambda_{res3}(\text{air})}{60}$$

8. Calculate D by solving Eq. 16-12:

$$D = \frac{c}{\sqrt{\epsilon_{eff}} f_{res2}} - 2(B + 2\Delta B - E)$$

9. Select C such that

$$\frac{C}{A} \geq 0.3 \quad \text{and} \quad \frac{C}{D} \geq 0.75$$

10. Calculate the effective permittivity and effective length extension of the pseudo patch of the fourth resonance with the effective width as $D - 2F$:

$$\epsilon_{eff(pp)} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(1 + \frac{12T}{D - 2F} \right)^{-1/2}$$

$$2\Delta_{B-E-F} = 0.824T \frac{(\epsilon_{eff(pp)} + 0.3) \left(\frac{D - 2F}{T} + 0.262 \right)}{(\epsilon_{eff(pp)} - 0.258) \left(\frac{D - 2F}{T} + 0.813 \right)}$$

11. Calculate H by solving Eq. 16-14:

$$H \approx B - E + 2\Delta_{B-E-H} - \frac{1}{\sqrt{\epsilon_{eff(pp)}}} \left(\frac{c}{f_{res4}} - (2C + D) \right)$$

12. Check that the sum $C + E + H$ is less than B . If not, adjust C by changing the ratios in step 9 and the value of H until the design is physically realizable.

16.9 VARIATIONS OF U-SLOT PATCH ANTENNA

Edge-Slotted v-Slot Antenna: The V-slot antenna was studied in Rafi and Shafai,⁴⁴ which showed that the V angle can be used to improve its performance. The bandwidth of a U -slot

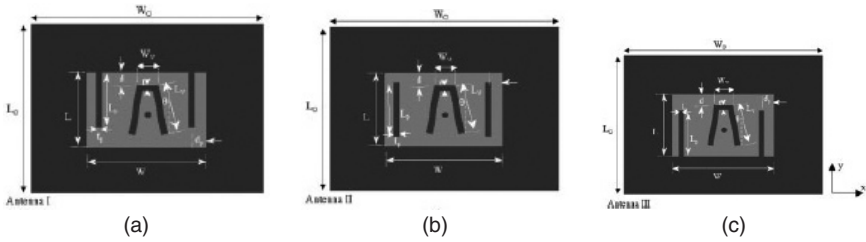


FIGURE 16-45 Geometries of wideband V-slotted microstrip patch antennas with two parallel-edge slots of different configurations: $\epsilon_r = 1.05$, $h = 5.5$ mm, $L_G = W_G = 60$ mm, $L = 26$ mm, $W = 36$ mm, $\theta = 10^\circ$, $L_V = 2.7$ mm, $W_V = 8.8$ mm, $t = 2$ mm, $d = 2.3$ mm, $L_p = 22$ mm, $t_p = 1.5$ mm, $d_p = 3$ mm⁴⁵

antenna was increased from about 30 percent to 36.5 percent by replacing the U -slot with a V -slot. The V angle is quite small and around 10° . A major problem with both U -slot and V -slot antennas is the cross polarization in the H -plane that increases rapidly at high frequencies. This problem was investigated and resolved in Rafi and Shafai.⁴⁵ Two additional slots were incorporated at the edge of the patch, as shown in Figure 16-45, which shows three different possibilities. Their return loss plots are shown in Figure 16-46. Antenna I shows a wideband performance similar to the original V -slot antenna without the edge slots. The cross polarizations in the H -plane are shown in Figure 16-47. For the original antenna, it increases rapidly beyond 5 GHz. However, for Antenna I, which has edge slots open at the top, the cross polarization is low and nearly independent of frequency. To investigate the effect of this cross-polarization improvement, its gain performance is compared with the original antenna in Figure 16-48. It is clear that the gain of the edge slotted antenna has increased by as much as 3 dB at high frequencies. This indicates that the radiated power of the cross-polar component has returned to the copolar component. Figure 16-49 shows the principal plane radiation patterns of the original and edge slotted Antenna I. The improvements in gain and cross polarization are evident.

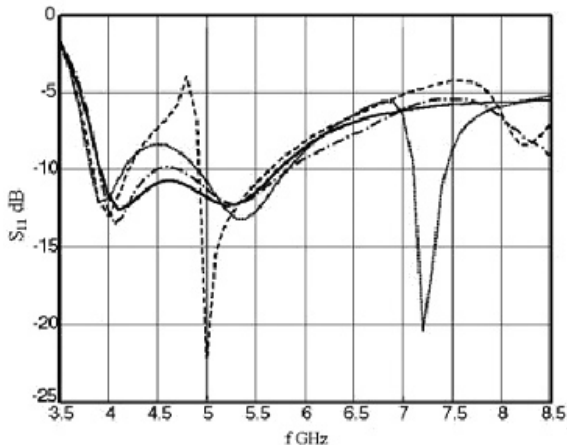


FIGURE 16-46 Return losses for the V -slot microstrip patch antennas of Figure 16-45: - - - - Without parallel slots; — With two parallel edge slots open at the top (Antenna I); ····· With two parallel edge slots at the middle (Antenna II); - · - · With two parallel edge slots open at the bottom (Antenna III) (after G. H. Rafi and L. Shafai⁴⁵ © *Microwave and Optical Technology Letters* 2004)

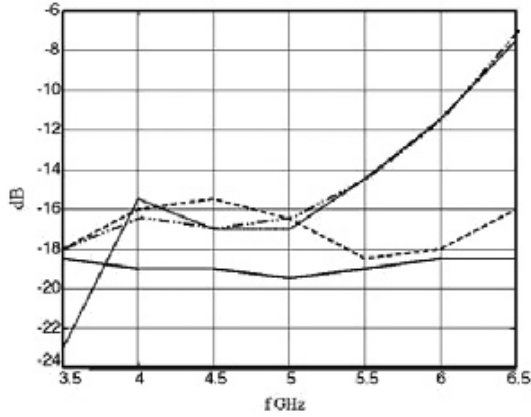


FIGURE 16-47 Cross-polarization levels in H -plane vs. frequency for different V -slot antennas of Figure 16-45: --- Without parallel slots; — With two parallel edge slots open at the top (Antenna I); ---- With two parallel edge slots at the middle (Antenna II); With two parallel edge slots open at the bottom (Antenna III) (after G. H. Rafi and L. Shafai⁴⁵ © *Microwave and Optical Technology Letters* 2004)

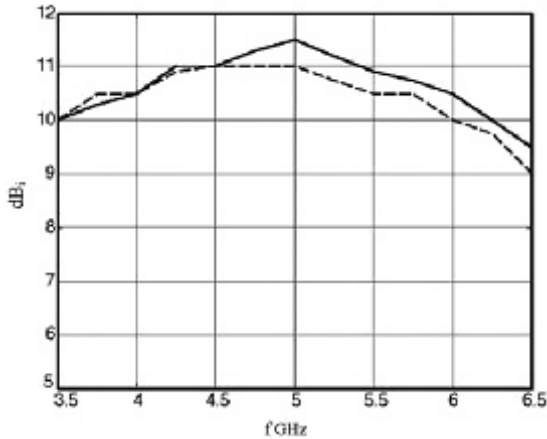


FIGURE 16-48 Available gain of V -slotted microstrip patch antenna in Figure 16-45: — With parallel slots; ---- Without parallel slots (Antenna I) (after G. H. Rafi and L. Shafai⁴⁵ © *Microwave and Optical Technology Letters* 2004)

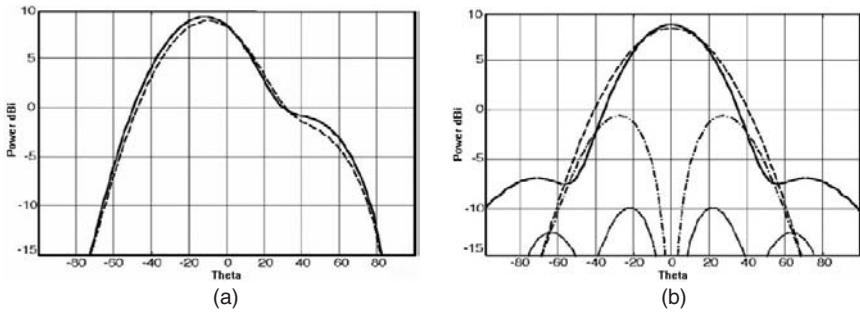


FIGURE 16-49 Radiation patterns and cross polarization of V-slotted antenna in Figure 16-45 at 6.5 GHz in the (a) *E*-plane and (b) *H*-plane: -.-.-.- Copol. V-slot; ——— Copol. V-slot with parallel slots (Antenna I); - - - - - Cross-pol. V-slot; Cross-pol. V-slot with parallel slots (Antenna I) (after G. H. Rafi and L. Shafai⁴⁵ © Microwave and Optical Technology Letters 2004).

Circular-Arc V-Slot Antenna: The patch shape in this antenna is changed from a rectangle to a double arc, as shown in Figure 16-50.⁴⁶ The bandwidth of this new antenna is increased to 40 percent, from 30 percent of the *U*-slot antenna. Other parameters remain similar.

Diamond-Shaped V-Slot Antenna: The patch in this antenna is diamond shaped, as shown in Figure 16-51.⁴⁷ Its return-loss plot in Figure 16-52 and the input impedance in Figure 16-53 show that there is an additional resonance, and the bandwidth increases to 50 percent.

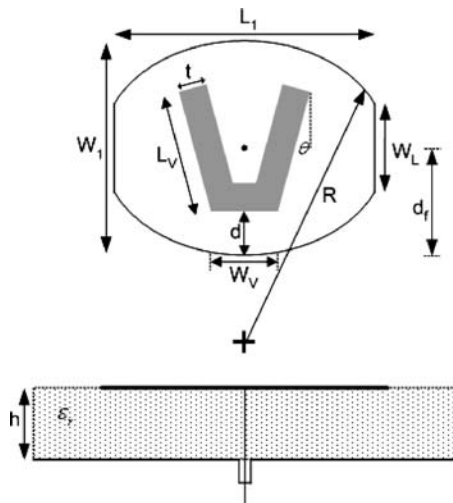


FIGURE 16-50 Geometry of the nonrectangular microstrip antenna with V-slot (after G. H. Rafi and L. Shafai⁴⁶ © IEEE 2003). The antenna parameters are $\epsilon_r = 1.0006$, $L_1 = 44$ mm, $W_1 = 27.4$ mm, $W_L = 8$ mm, $R = 29.2$ mm, $\theta = 7^\circ$, $L_V = 20$ mm, $W_V = 8.8$ mm, $t = 2$ mm, $d = 2$ mm, $d_1 = 18$ mm (after Rafi and Shafai⁴⁷ © Electronics Letters 2004).

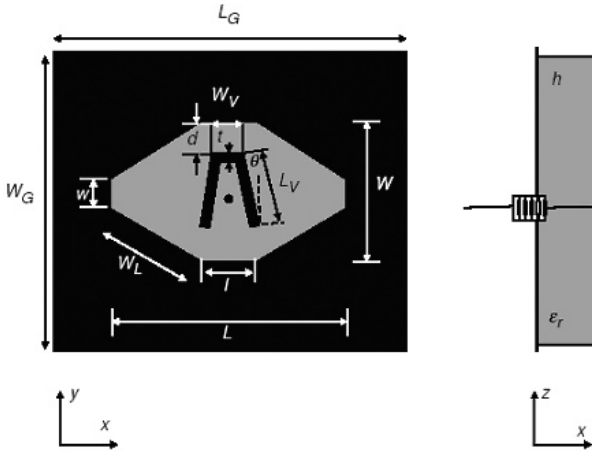


FIGURE 16-51 Geometry of wideband microstrip patch antennas with V-slot and diamond-shaped patch: $L_G = 100$ mm, $W_G = 80$ mm, $h = 5$ mm, $\epsilon_r = 1.05$, $L = 58$ mm, $W = 32$ mm, $W_L = 28$ mm, $w = 8$ mm, $t = 8$ mm, $\theta = 12.5^\circ$, $L_V = 19$ mm, $W_V = 8.8$ mm, $d = 4.3$ mm (after Rafi and Shafai⁴⁷ © *Electronics Letters* 2004)

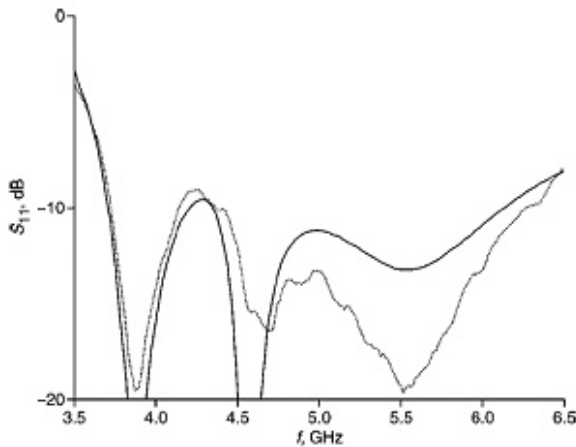


FIGURE 16-52 Comparison of simulated and measured return losses of V-slot microstrip antenna with diamond-shaped patch and data as in Figure 16-51 (after Rafi and Shafai⁴⁷ © *Electronics Letters* 2004).

..... Measured ——— Calculated

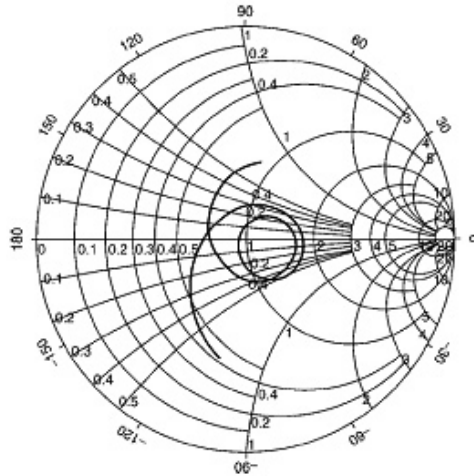


FIGURE 16-53 Impedance plot for V-slotted diamond-shaped patch antenna and data same as in Figure 16-51 (after Rafi and Shafai⁴⁷ © *Electronics Letters* 2004)

Double U-Slot Antenna: In this antenna a second U-slot is placed inside the first one⁴⁸ (see Figure 16-54). Its bandwidth has increased to 44 percent (1.24–1.95 GHz), as shown in Figure 16-55. The new resonance is evident from the second loop on the Smith Chart of Figure 16-56.

Circular Patch L-Probe Antenna: An L-probe is used to feed a circular patch with a circular-slot, as shown in Figure 16-57.⁴⁹ Its bandwidth is 40 percent, from 3.55 to 5.32 GHz, as shown in Figure 16-58. Within the bandwidth, its gain was above 8 dBi.

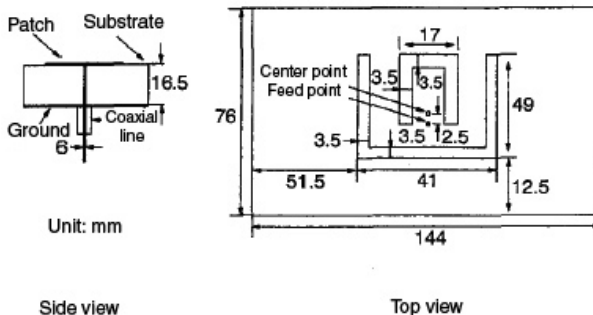


FIGURE 16-54 Geometry of double U-slot patch antenna on a foam substrate (Y. X. Guo et al⁴⁸ © *Electronics Letters* 1998)

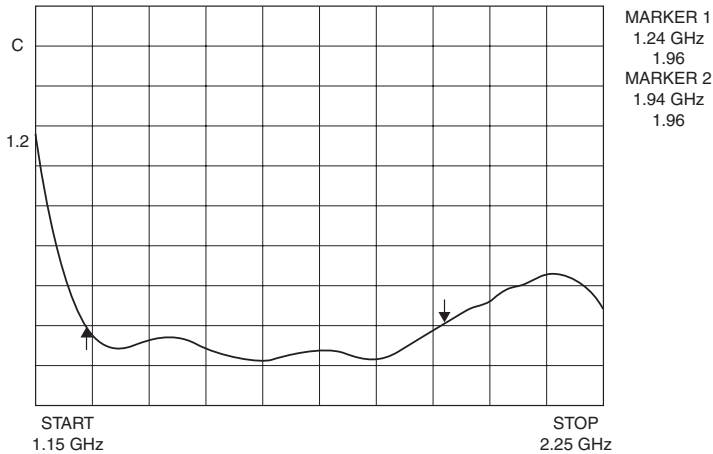


FIGURE 16-55 SWR against frequency of the antenna with double *U*-slot in Figure 16-54 (Y. X. Guo et al⁴⁸ © *Electronics Letters* 1998)

Stacked slotted antennas have also been investigated for achieving even larger bandwidths. In Clénet et al⁵⁰ a *U*-slotted patch is stacked with a conventional patch, and bandwidth of 45 percent is obtained. In Rafi and Shafai⁵¹ two *U*-slotted patches are stacked and investigated (see Figure 16-59). An impedance bandwidth of 46 percent is obtained. In Rafi and Shafai⁵² a *V*-slotted patch antenna is first stacked with a conventional patch (see Figure 16-60) and obtained a bandwidth of 47 percent. Then it is stacked with another *V*-slotted patch (see Figure 16-61). The bandwidth of the new antenna was increased to 54 percent.

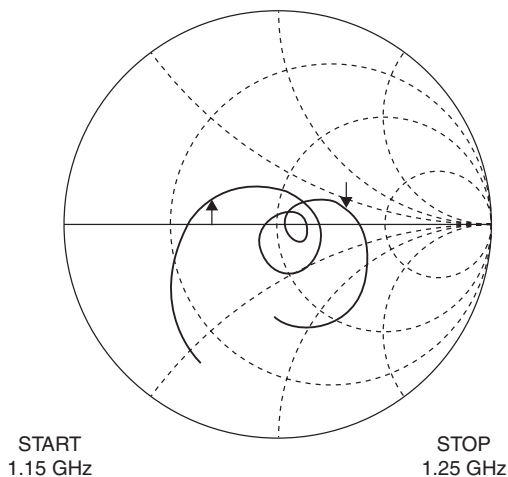


FIGURE 16-56 Input impedance of the antenna with double *U*-slot in Figure 16-54 (Y. X. Guo et al⁴⁸ © *Electronics Letters* 1998)

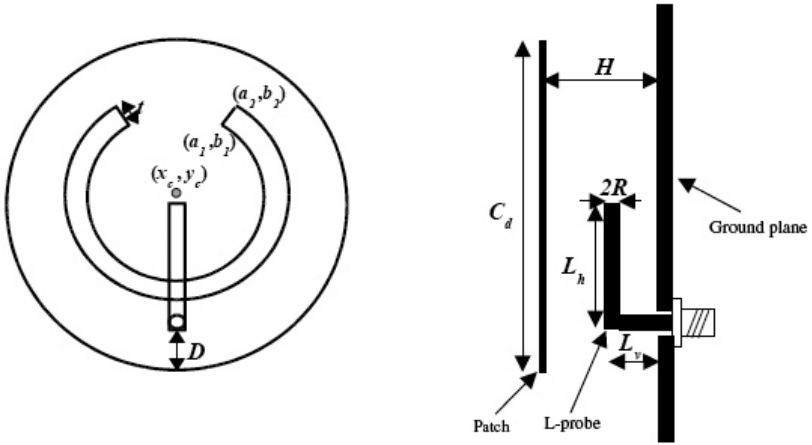


FIGURE 16-57 Geometry of a circular patch antenna with a circular arc slot and an L -shaped probe⁴⁹

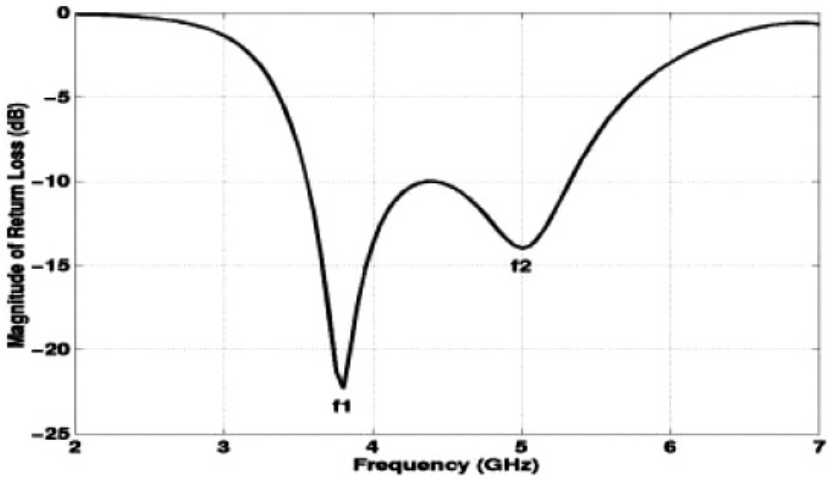


FIGURE 16-58 Return loss of a circular patch antenna with a circular arc slot and an L -shaped probe:⁴⁹ $C_d = 34$ mm, $(x_c, y_c) = (0$ mm, 1.75 mm), $(a_1, b_1) = (5$ mm, 8.75 mm), $(a_2, b_2) = (7$ mm, 10.75 mm), $t = 2.83$ mm, $L_h = 14$ mm, $L_v = 5.5$ mm, $R = 1$ mm, $D = 3$ mm, foam substrate height $H = 9$ mm

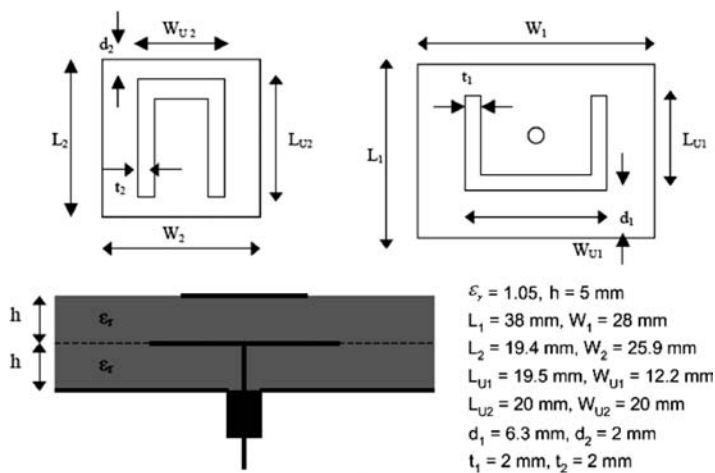
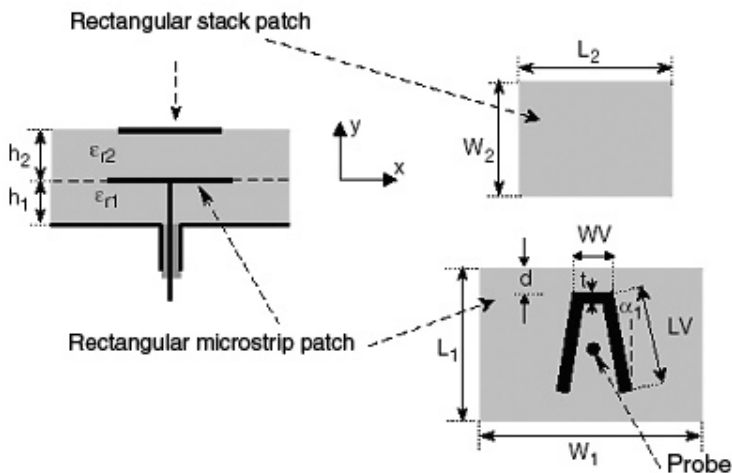
FIGURE 16-59 Stacked U-slot antenna⁵¹

FIGURE 16-60 Geometry of a double-layer V-slotted rectangular patch stacked with simple rectangular patch: $\epsilon_{r1} = \epsilon_{r2} = 1.05$, $h_1 = h_2 = 5$ mm, $L_1 = 37$ mm, $W_1 = 28$ mm, $\alpha_1 = 10^\circ$, $L_v = 19$ mm, $W_v = 7.2$ mm, $d_1 = 6.3$ mm, $t = 2$ mm, $L_2 = 26$ mm, $W_2 = 19.4$ mm (after G. H. Rafi and L. Shafai⁵² © IEE 2002)

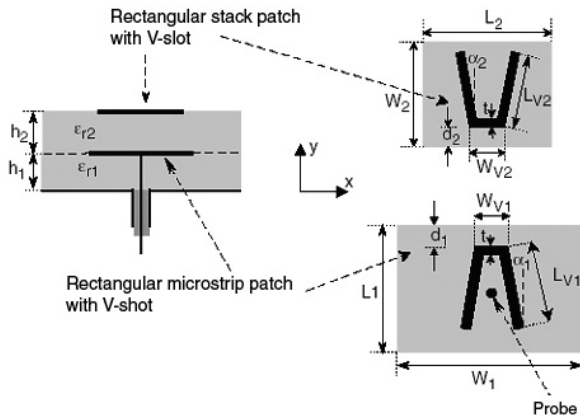


FIGURE 16-61 Geometry of double-layer V-slotted rectangular microstrip patch stacked with another V-slotted patch: $\epsilon_{r1} = \epsilon_{r2} = 1.0006$, $h_1 = h_2 = 5$ mm, $L_1 = 37$ mm, $W_1 = 28$ mm, $\alpha_1 = 7^\circ$, $L_{V1} = 19$ mm, $W_{V1} = 8$ mm, $d_1 = 6.3$ mm, $t = 2$ mm, $L_2 = 26$ mm, $W_2 = 19.4$ mm, $\alpha_2 = 24^\circ$, $L_{V2} = 18$ mm, $W_{V2} = 6$ mm, $d_2 = 1.7$ mm (after G. H. Rafi and L. Shafai⁵² © IEE 2002)

16.10 EBG LOADED SLOTTED PATCH ANTENNAS

Electromagnetic band gap structures interact strongly with the substrate field and have been used to reduce surface wave excitation and losses in microstrip antennas. In this case an attempt was made to use them for further increasing the patch bandwidth. The results have been surprisingly good. In Rafi and Shafai⁵³ small shorted patches were placed parallel to the antenna radiating edges to intercept the substrate mode fields. They generate a high impedance barrier to the substrate fields and reradiate their power, improving the bandwidth and gain. The resulting antenna geometry is shown in Figure 16-62, and its return

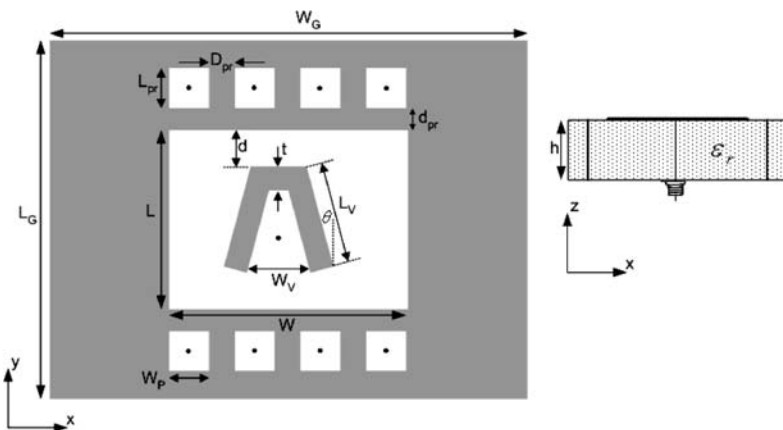


FIGURE 16-62 Geometry of the wideband microstrip patch antennas with V-slot and patch-via resonators: $L = 26$ mm, $W = 35$ mm, $L_G = 52$ mm, $W_G = 80$ mm, $\theta = 16.5^\circ$, $L_V = 20.7$ mm, $W_v = 8.8$ mm, $d = 2.3$ mm, $\epsilon_r = 1.0006$, $h = 5.5$ mm, $t = 2$ mm, $d_{pr} = 4$ mm, $D_{pr} = 5$ mm, $L_{pr} = 5$ mm (after G. H. Rafi and L. Shafai⁵³ © IEEE 2004)

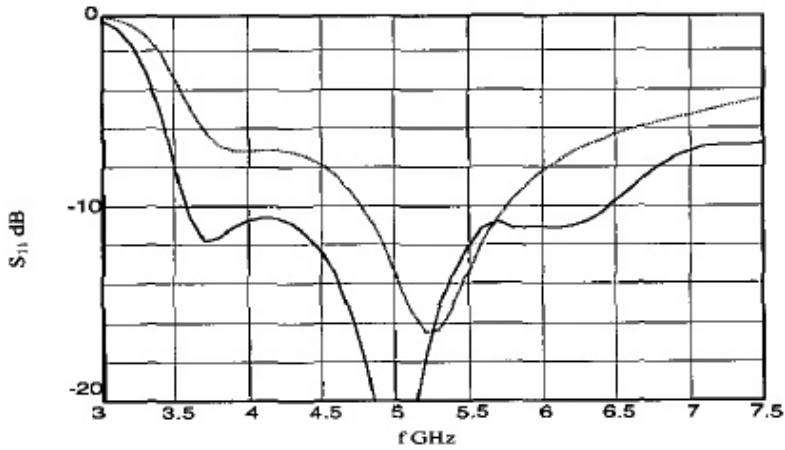


FIGURE 16-63 Comparison of return loss for V-slot microstrip patch with and without patch-via resonators: V-slot microstrip patch without patch-via resonators. ——— V-slot microstrip patch with patch-via resonators (after G. H. Rafi and L. Shafai⁵³ ©IEEE 2004).

loss in Figure 16-63. An impedance bandwidth of 58 percent was obtained. This is significant, since it is obtained with a single-layer design. Its radiation characteristics remained similar to the original antenna.

A similar approach was also used for loading a slotted-diamond patch antenna.⁵⁴ The loading EBG structures, however, were two periodically shorted narrow patches, in lieu of arrays of small EBG patches. The antenna geometry is shown in Figure 16-64. The VSWR plots for both rectangular and diamond-shaped patches are shown in Figure 16-65. The achieved bandwidth of the latter antenna was 66 percent, much higher than that of

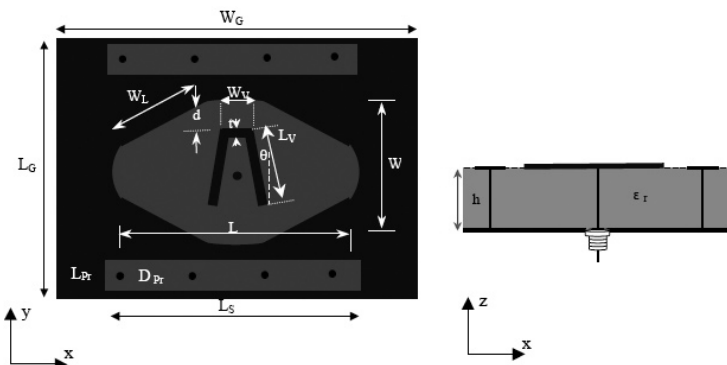


FIGURE 16-64 Geometry of the wideband diamond-shaped patch antennas with V-slot and strip-via resonators:⁵⁴ $L = 46$ mm, $W = 28$ mm, $L_G = 52$ mm, $W_G = 80$ mm, $\theta = 12.5^\circ$, $L_V = 19$ mm, $W_V = 8.8$ mm, $d = 4.3$ mm, $h = 5$ mm, $\epsilon_r = 1.1$, $D_{pr} = 10$ mm, $L_{pr} = 6$ mm, $t = 2$ mm

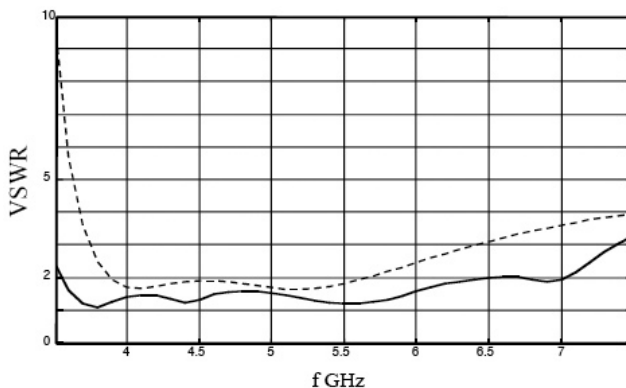


FIGURE 16-65 VSWR of wideband diamond-shaped patch antenna with strip-via resonators, Figure 16-64, and original V-slotted rectangular patch microstrip patch antenna.⁵⁴
 Rectangular patch ——— Diamond shape patch

even stacked patches. Figure 16-66 shows the gains and cross polarizations. The gain of the loaded antenna is enhanced, especially at high frequencies, making it independent of frequency. Because in slotted patch antennas the gain decreases rapidly at high frequencies, the improvement in the gain is also remarkable. The midband radiation patterns are shown in Figure 16-67.

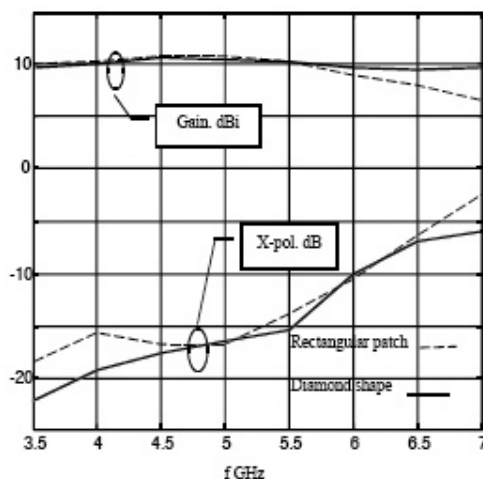


FIGURE 16-66 Antenna gain and cross polarization (X-pol.) vs. frequency for diamond-shaped patch antenna with strip-via, Figure 16-64, and original V-slotted rectangular patch⁵⁴

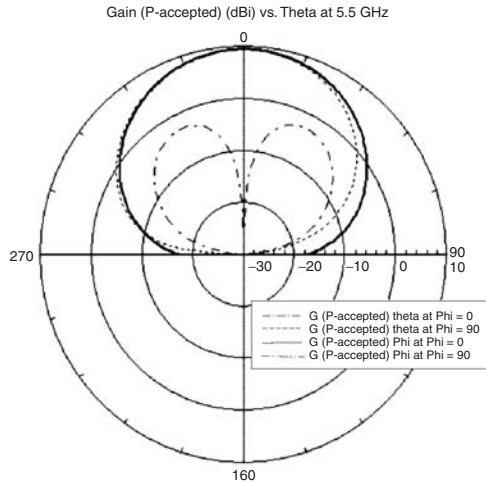


FIGURE 16-67 Antenna E - and H -plane ($\phi = 0$, $\phi = 90$) patterns at 5.5 GHz; data same as in Figure 16-64⁵⁴

16.11 MICROSTRIP PATCH ANTENNAS WITH EBG SUBSTRATE

In microstrip patch and other low-profile antennas over a conducting ground plane, the negative image of the radiating electric current detrimentally influences its radiation efficiency and bandwidth. In contrast, a magnetic ground plane would enhance the radiation, and thus the efficiency and bandwidth. However, since a perfect magnetic ground plane does not exist, it should be developed using periodic materials such as EBG structures, which have finite band gaps. As a result, the bandwidth of a so-generated magnetic ground plane will also be finite, and restricted to its band gap, outside which it will have a complex surface impedance similar to tuned circuits. In this section the properties of a microstrip patch antenna over a high-impedance EBG substrate is investigated within its band gap.

The geometry of the antenna is shown in Figure 16-68. A rectangular patch antenna is placed over a high-impedance EBG whose band gap occurs within a frequency band of 5.8–7.5 GHz.⁵⁵ A reference patch antenna, i.e., the same rectangular patch over a dielectric substrate of similar thickness (1.59 mm) and size with a perfect electric ground plane, was also designed for operation at 6.6 GHz.

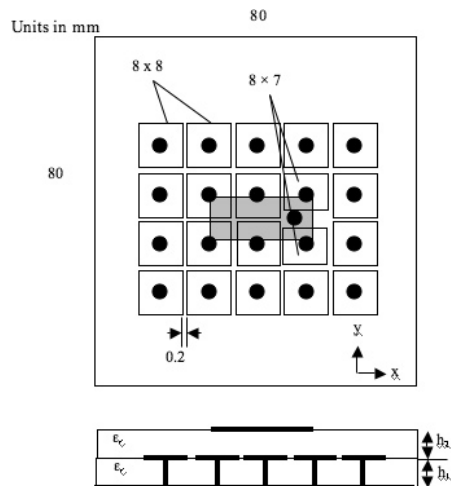


FIGURE 16-68 Geometry of a microstrip patch with an EBG substrate, patch length $L = 14$ mm, width $W = 6$ mm, $\epsilon_r = 2.5$, $h_1 = h_2 = 1.59$ mm, feed distance from patch center $L_f = 5.2$ mm (after D. Qu et al⁵⁵ © IEE 2006)

TABLE 16-2 Minimum S_{11} , Bandwidth, and Gain for the Antenna in Figure 16-68 vs. the Feed Location L_f

Feed Location L_f (mm)	Minimum S_{11} (dB)	Impedance Bandwidth	Gain at $\theta = 0^\circ$ (dBi)
4.6	-22.44	19.91% (6.24–7.62 GHz)	10.31
5.2	-44.77	20.20% (6.23–7.63 GHz)	10.32
5.8	-35.75	16.38% (6.50–7.66 GHz)	10.32

Because of the small size of the ground plane (80 mm \times 80 mm), its performance in gain was poor and provided a boresight gain of only 3.92 dBi. Its bandwidth was also small and about 3.85 percent. The microstrip patch over the EBG substrate was also designed for 6.6 GHz and had the parameters Length $L = 14$ mm, Width $W = 6$ mm, $\epsilon_r = 2.5$, $\tan \delta = 0.001$, substrate heights $h_1 = h_2 = 1.59$ mm, and probe distance from the patch center $L_f = 5.2$ mm. The two central EBG patches are 8 mm \times 7 mm, to accommodate the feed probe. Other EBG patches are 8 mm \times 8 mm, with a spacing of 0.2 mm. The ground plane size is 80 mm \times 80 mm.

A parametric study was conducted for this antenna. First, the feed location L_f was assigned different values. Its effects, which are significant, on the patch minimum S_{11} , bandwidth, and gain are shown in Table 16-2. When $L_f = 5.2$ mm, the impedance bandwidth reaches 20.20 percent (6.23–7.63 GHz), and minimum S_{11} is -44.77 dB. The gain is not sensitive to the feed location, and its maximum of 10.32 dBi is obtained for this feed location, as the mismatch loss is minimal. The gain bandwidth was larger than its impedance bandwidth and remained within 1.0 dB of its peak in about 15 percent bandwidth. Next, the effect of patch width was studied. The results are shown in Table 16-3, again showing significant effects on the impedance bandwidth. The gain bandwidth was large similar to the previous case.

For this antenna, with $W = 6$ mm, $L_f = 5.2$ mm, effects of different ground plane sizes were also studied. The results are shown in Table 16-4, indicating a relatively minor effect on the patch bandwidth. The selected ground plane sizes are around 1 to 1.5 wavelengths, where the patch gain and bandwidth remained nearly constant, larger than that of the impedance bandwidth. The gain and its bandwidth remained large, as long as the ground plane size was larger than one-half wavelength.

TABLE 16-3 Minimum S_{11} , Bandwidth, and Gain for the Antenna of Figure 16-68 vs. the Patch Width

Patch Width (mm)	Minimum S_{11} (dB)	Impedance Bandwidth	Gain at $\theta = 0^\circ$ (dBi)
5.5	-33.93	16.12% (6.50–7.64 GHz)	10.34
6.0	-44.77	20.20% (6.23–7.63 GHz)	10.32
6.5	-23.42	16.34% (6.35–7.48 GHz)	10.32

TABLE 16-4 Minimum S_{11} , Bandwidth, and Gain for the Antenna of Figure 16-68 vs. Its Ground Plane Size

Ground Plane Size $L \times W$ (mm)	Minimum S_{11} (dB)	Impedance Bandwidth	Gain at $\theta = 0^\circ$ (dBi)
55×55	-29.75	18.64% (6.42–7.74 GHz)	10.20
65×65	-26.62	18.13% (6.42–7.70 GHz)	10.39
80×80	-44.77	20.20% (6.23–7.63 GHz)	10.32

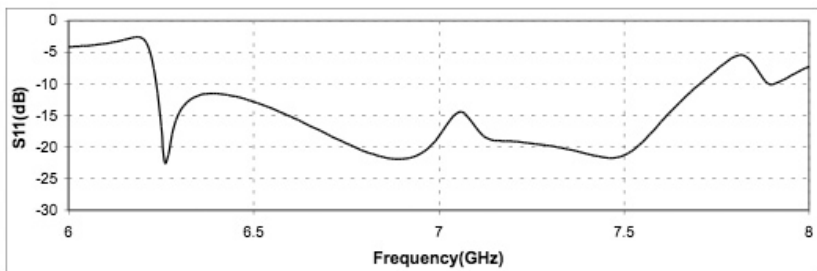
TABLE 16-5 Minimum S_{11} , Bandwidth, and Gain for the Antenna of Figure 16-68 vs. the EBG Spacing

EBG Spacing (mm)	Minimum S_{11} (dB)	Impedance Bandwidth	Gain at $\theta = 0^\circ$ (dBi)
0.1	-34.42	17.97% (6.23–7.46 GHz)	9.67
0.2	-44.77	20.20% (6.23–7.63 GHz)	10.32
0.3	-22.41	21.08% (6.24–7.71 GHz)	9.81

For the antenna with $80 \text{ mm} \times 80 \text{ mm}$ ground plane, the effect of EBG patch spacing was studied next. The results are shown in Table 16-5, affecting all antenna parameters. The impedance bandwidth is broadest (21.08 percent) when the EBG spacing is 0.3 mm. Its S_{11} plot is shown in Figure 16-69.

The radiation patterns of the antenna of Figure 16-68 with the EBG substrate are shown in Figures 16-70*a* and 16-70*b*, at 6.26 GHz. The maximum gain is 9.81 dBi at $\theta = 0^\circ$, which is much higher than that of the reference patch (3.92 dBi). The H -plane cross polarization has a peak level of -21 dB.

The results in this study showed that a microstrip patch antenna over an EBG ground plane and within its band gap has a significantly improved performance in its impedance bandwidth, gain, and cross polarization. In addition, the patch gain was relatively insensitive to its parameters and remained high within its impedance bandwidth. The improvements in gain and bandwidth are expected from antennas over a magnetic ground plane, and the selected EBG substrate behaves similarly within its band gap.

**FIGURE 16-69** The S_{11} of microstrip antenna of Figure 16-68 with EBG patch spacing of 0.3 mm (after D. Qu et al.⁵⁵ © IEE 2006)

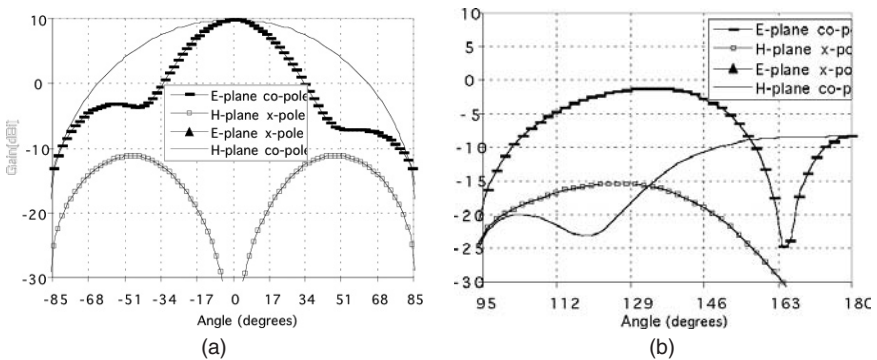


FIGURE 16-70 E- and H-plane radiation patterns of microstrip antenna of Figure 16-68 with EBG patch spacing of 0.3 mm, at 6.26 GHz: (a) front radiation and (b) back radiation (after D. Qu et al¹⁵⁵ © IEE 2006)

REFERENCES

1. J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna Theory and Design* (London: Peter Peregrinus, 1981).
2. J. F. Zurcher and F. E. Gardiol, *Broadband Patch Antennas* (Norwood, Mass.: Artech House, 1995).
3. G. Kumar and K. P. Ray, *Broadband Microstrip Antennas* (Norwood, Mass.: Artech House, 2003).
4. K.-L. Wong, *Compact Broadband Microstrip Antennas* (New York: John Wiley & Sons, Inc., 2002).
5. D. M. Pozar and D. H. Schaubert, "A Review of Bandwidth Enhancement Techniques for Microstrip Antennas," *Microstrip Antennas, The Analysis and Design of Microstrip Antennas and Arrays*, (eds.) (New York: IEEE Press, 1995): 157–166.
6. D. Sanchez-Hernandez and I. D. Robertson, "A Survey of Broadband Microstrip Patch Antennas," *Microwave Journal* (September 1996): 60–84.
7. D. M. Pozar, "Microstrip Antennas," *Proc. IEEE*, vol. 80 (January 1992): 79–91.
8. J. R. James and P. S. Hall, *Handbook of Microstrip Antennas*, vol. 1 (London: Peter Peregrinus, 1989).
9. P. M. Bafrooei and L. Shafai, "Characteristics of Single- and Double-Layer Microstrip Square Ring Antennas," *IEEE Trans. Antennas Propagat.*, vol. 47, no. 10 (October 1999): 1633–1639.
10. H. F. Pues and A. R. Van de Capelle, "An Impedance Matching Technique for Increasing the Bandwidth of Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 37 (November 1989): 1345–1354.
11. D. M. Pozar and B. Kaufman, "Increasing the Bandwidth of a Microstrip Antenna by Proximity Coupling," *Electronics Letters*, vol. 23, no. 8 (April 1987): 368–369.
12. G. Splitt and M. Davidovitz, "Guidelines for Design of Electromagnetically Coupled Microstrip Patch Antennas on Two-Layer Substrates," *IEEE Trans. Antennas Propagat.*, vol. 38, no. 7 (July 1990): 1136–1140.
13. B. Lethakumary, S. K. Menon, C. K. Anandan, K. Vasudevan, and P. Mohanan, "L-Strip Excited Wideband Rectangular Microstrip Antenna," *Microwave and Optical Technology Letters*, vol. 42, no. 2 (July 2004): 173–175.
14. B. Lethakumary, S. K. Menon, P. Francis, C. K. Anandan, K. Vasudevan, and P. Mohanan, "Wideband Microstrip Antenna Using Hook-Shape Feed," *Microwave and Optical Technology Letters*, vol. 44, no. 2 (January 2005): 169–171.

15. S. K. Menon, B. Lethakumary, K. Vasudevan, and P. Mohanan, "Wide Band Rectangular Microstrip Antenna Using Symmetric T-Shaped Feed," *Microwave and Optical Technology Letters*, vol. 35, no. 3 (November 2002): 235–236.
16. H. Legay and L. Shafai, "Parametric Analysis of an EMC Patch Surrounded by Parasitic Elements," *AEU*, vol. 48, no. 4 (1994): 184–190.
17. D. M. Pozar, "Microstrip Antenna Aperture Coupled to a Microstrip Line," *Electronics Letters*, vol. 21, no. 2 (January 1985): 49–50.
18. J. F. Zurcher, "The SSFIP: A Global Concept for High Performance Broadband Planar Antennas," *Electronics Letters*, vol. 24, no. 13 (June 1988): 1433–1435.
19. D. M. Pozar and S. D. Targonski, "Improved Coupling for Aperture Coupled Microstrip Antennas," *Electronics Letters*, vol. 27, no. 13 (June 1991): 1129–1131.
20. X. H. Yang and L. Shafai, "Characteristics of Aperture-Coupled Microstrip Antennas with Various Radiating Patches and Coupling Slots," *IEEE Trans. Antennas Propagat.*, vol. 43, no. 1 (January 1995): 72–78.
21. V. Rathi, G. Kumar, and K. P. Ray, "Improved Coupling for Aperture Coupled Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 44, no. 8 (August 1996): 1196–1198.
22. F. Croq and A. Papiernik, "Large Bandwidth Aperture-Coupled Microstrip Antenna," *Electronics Letters*, vol. 26, no. 16 (August 1990): 1293–1294.
23. L. Giauffret, J. M. Laheurte, and A. Papiernik, "Experimental and Theoretical Investigation of New Compact Large Bandwidth Aperture Coupled Microstrip Antenna," *Electronics Letters*, vol. 31, no. 25 (December 1995): 2139–2140.
24. Y. Qin, S. Gao, A. Sambeli, M. Edison, and E. Korolkiewicz, "Design of a Broadband Square-Ring-Slot Couple Patch Antenna," *Microwave and Optical Technology Letters*, vol. 47, no. 5 (December 2005): 454–457.
25. Ansoft Ensemble Version 8, Ansoft Corporation, USA.
26. H. Legay and L. Shafai, "New Stacked Microstrip Antenna with Large Bandwidth and High Gain," *IEE Proc. Microwave, Antennas and Propagation*, Pt. H, vol. 141, no. 3 (June 1994): 199–204.
27. A. Sabban, "A New Broadband Stacked Two Layer Microstrip Antenna," *IEEE AP-S Int. Symposium Digest* (June 1983): 63–66.
28. G. Kumar and K. C. Gupta, "Broadband Microstrip Antennas Using Additional Resonances Gap-Coupled to the Radiating Edges," *IEEE Trans. Antennas Propagat.*, vol. 32, no. 12 (December 1984): 1375–1379.
29. F. Croq, G. Kossivas, and A. Papiernik, "Stacked Resonators for Bandwidth Enhancement: A Comparison of Two Feeding Techniques," *IEE Proc. Microwave, Antennas and Propagation*, Pt. H, vol. 140, no. 4 (August 1993): 303–308.
30. R. B. Waterhouse, "Design of Probe-Fed Stacked Patches," *IEEE Trans. Antennas Propagat.*, vol. 47, no. 12 (December 1999): 1780–1784.
31. W. S. T. Rowe and R. B. Waterhouse, "Broadband CPW Fed Stacked Patch Antenna," *Electronics Letters*, vol. 35, no. 9 (April 1999): 681–682.
32. S. D. Targonski, R. B. Waterhouse, and D. M. Pozar, "Design of Wide-Band Aperture-Stacked Patch Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 46, no. 9 (September 1998): 1245–1251.
33. C. L. Mak, K. M. Luk, K. F. Lee, and Y. L. Chow, "Experimental Study of a Microstrip Antenna with an L-Shaped Probe," *IEEE Trans. Antennas Propagat.*, vol. 48, no. 5 (May 2000): 777–783.
34. K. M. Luk, C. L. Mak, Y. L. Chow, and K. F. Lee, "Broadband Microstrip Patch Antenna," *Electronics Letters*, vol. 34, no. 15 (July 1998): 1442–1443.
35. P. Li, K. L. Lau, K. M. Luk, "A Study of Wide-Band L-Probe Fed Planar Patch Antenna Mounted on a Cylindrical or Conical Surface," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 10 (October 2005): 3385–3389.
36. K. M. Luk, L. K. Au Yeung, C. L. Mak, and K. F. Lee, "Circular Patch Antenna with an L-Shaped Probe," *Microwave and Optical Technology Letters*, vol. 20, no. 4 (February 1999): 256–257.

37. C. L. Mak, K. F. Lee, and K. M. Luk, "Broadband Patch Antenna with a T-Shaped Probe," *IEE Proc. Microwave, Antennas and Propagation*, Pt. H, vol. 147, no. 2 (April 2000): 73–76.
38. T. Huynh and K. F. Lee, "Single-Layer Single-Patch Wideband Microstrip Antenna," *Electronics Letters*, vol. 31, no. 16 (August 1995): 1310–1312.
39. M. Clénet and L. Shafai, "Multiple Resonances and Polarization of U-Slot Patch Antenna," *Electronics Letters*, vol. 35, no. 2 (June 1999): 101–102.
40. K. F. Lee, K. M. Luk, K. F. Tong, S. M. Shum, T. Huynh, and R. Q. Lee, "Experimental and Simulation Studies of the Coaxially Fed U-Slot Rectangular Patch Antenna," *IEE Proc. Microwave, Antennas and Propagation*, Pt. H, vol. 144, no. 5 (October 1997): 354–358.
41. R. Bhalla and L. Shafai, "Resonance Behavior of Single U-Slot Microstrip Patch Antenna," *Microwave and Optical Technology Letters*, vol. 32, no. 5 (March 2002): 333–335.
42. R. Bhalla and L. Shafai, "Broadband Patch Antenna with a Circular Arc Shaped Slot," *IEEE AP-S Int. Symposium*, vol. 1 (June 2002): 394–397.
43. S. Weigand, G. H. Huff, K. H. Pan, and J. T. Bernhard, "Analysis and Design of Broad-Band Single-Layer Rectangular U-Slot Microstrip Patch Antennas," *IEEE Trans. Antennas Propagat.*, vol. 51, no. 3 (March 2003): 457–468.
44. G. H. Rafi and L. Shafai, "V-Slot Antennas for Wideband Applications," *Symposium on Antenna Technology and Applied Electromagnetics, ANTEM 2002* (July 2002): 463–466.
45. G. H. Rafi and L. Shafai, "Low Cross-Polarized Wideband V-Slot Microstrip Antenna," *Microwave and Optical Technology Letters*, vol. 43, no. 1 (October 2004): 44–47.
46. G. H. Rafi and L. Shafai, "Circular-Arc Single Layer Microstrip Antenna for Wideband Application," *IEEE AP-S Int. Symposium Digest*, vol. 2 (June 2003): 260–263.
47. G. H. Rafi and L. Shafai, "Wideband V-Slotted Diamond-Shaped Microstrip Antennas," *Electronics Letters*, vol. 40, no. 19 (September 2004): 1166–1167.
48. Y. X. Guo, K. M. Luk, K. F. Lee, and Y. L. Chow, "Double U-Slot Rectangular Patch Antenna," *Electronics Letters*, vol. 34, no. 19 (September 1998): 1805–1806.
49. R. Bhalla and L. Shafai, "Circular Patch Antenna with a Circular Arc-Slot and an L-Probe Feeding," *Symposium on Antenna Technology and Applied Electromagnetics, ANTEM 2002* (July 2002): 456–462.
50. M. Clénet, C. B. Ravipati, and L. Shafai, "Effect of Parasitic Patch on the Bandwidth Characteristics of U-Slot Rectangular Microstrip Antenna," *Symposium on Antenna Technology and Applied Electromagnetics, ANTEM'98* (August 1998): 229–232.
51. G. H. Rafi and L. Shafai, "The Effect of Stacked U-Slot Patch on the Radiation Parameters of Rectangular U-Slot Microstrip Antenna," *Symposium on Antenna Technology and Applied Electromagnetics, ANTEM 2002* (July 2002): 451–454.
52. G. H. Rafi and L. Shafai, "Broadband Microstrip Patch Antenna with V-Slot," *IEE Proc. Microwave, Antennas and Propagation*, vol. 151, no. 5 (October 2004): 435–440.
53. G. H. Rafi and L. Shafai, "Wideband Microstrip Patch Antennas with V-Slot and Patch-Via Resonators," *IEEE APS Int. Symposium Digest* (June 2004): 3741–3744.
54. G. H. Rafi and L. Shafai, "Improving Bandwidth of Single Layer Diamond Shape Patch Microstrip Antenna with Strip-Via Resonators," *Symposium on Antenna Technology and Applied Electromagnetics, ANTEM 2004* (July 2004): 673–676.
55. D. Qu, L. Shafai and A. Foroozesh, "Improving Microstrip Patch Antenna Performance Using EBG Substrates," *IEE Proc. Microwave, Antennas and Propagation*, Pt. H, vol. 153, no. 6 (December 2006): 558–563.

Chapter 17

Dielectric Resonator Antennas

Ahmed A. Kishk

*Professor, Department of Electrical Engineering,
Center for Applied Electromagnetic Systems Research
(CAESR), University of Mississippi*

Yahia M.M. Antar

*Canada Research Chair in Electromagnetic
Engineering, Professor, Department of Electrical
Engineering, Royal Military College of Canada*

CONTENTS

17.1	INTRODUCTION.....	17-2
17.2	POSSIBLE METHODS OF ANALYSIS	17-7
17.3	PROGRESS IN DRA RESEARCH.....	17-9
17.4	DRA ARRAYS.....	17-16

17.1 INTRODUCTION

History of Dielectric Resonator Antennas

The term *dielectric resonator* (DR) first appeared in 1939, when Ritchmyer¹ of Stanford University showed that unmetallized dielectric objects in the form of *toroids* could function as microwave resonators. However, his theoretical investigations failed to generate significant interest and practically nothing happened in this area for over 25 years. In the early 1960s, researchers from Columbia University, Okaya and Barash,^{2,3} reported the first ever DR in the form of a single crystal TiO_2 . In the mid-1960s, Cohn⁴ and his coworkers at Rantec Corporation performed the first extensive theoretical and experimental evaluation of DR. Nevertheless, the DR was still far from being used in practical applications. A real breakthrough in the dielectric ceramic industry occurred in the early 1970s when Masse et al developed the first temperature-stable, low-loss barium tetratitanate ceramics.⁵ In 1975, Van Bladel⁶ reported detailed theory to evaluate the modes of DR. The emergence of DR as a circuit element for several microwave wireless devices has injected tremendous interest into the research of such material.⁷⁻⁹ With the development of low-loss dielectrics, resonators with high Q elements for circuit applications such as filters and oscillators were developed, thus increasing the possibilities for developing more compact integrated microwave circuits.⁸ For these circuits, resonators, which are typically cylindrical in shape, are made out of high dielectric constant materials ($\epsilon_r > 35$) and are usually shielded to maintain the high quality factor needed for applications in oscillators, filters, etc.

DRs of different shapes have various modes of oscillation. With the proper excitation of certain modes and with no shielding, these resonators can actually become efficient radiators instead of energy storage devices. This concept led to the exploration of DRs as antennas. Although open DRs were found to radiate many years ago,^{10,11} the idea of using DR as a practical antenna materialized in 1983, when Long et al¹² published a paper on cylindrical dielectric resonator antennas (DRAs). Subsequently, other researchers investigated rectangular¹³ and hemispherical¹⁴ DRAs. These works laid the foundation for later extensive investigations on various aspects of DRAs. Considerable work done on analyzing their resonant modes, radiation characteristics, and various feeding schemes has demonstrated that these “new” radiators could offer new and attractive features in antenna design. Many new elements and arrays with attractive characteristics for wireless and other applications have been implemented and descriptions of some of these are included in the literature.¹⁵⁻¹⁷

Advantages of DRAs

As has been recently demonstrated, DRAs offer a high degree of flexibility and versatility over a wide frequency range, allowing for designers to suit many requirements. DRAs offer the following advantages:

- The DRA size is proportional to $\lambda_o/\epsilon_r^{1/2}$, where λ_o is the wavelength at resonant frequency and ϵ_r is the dielectric constant of the DR. Thus for the same frequency there is a natural reduction in size, compared with their conventional counterparts like microstrip antennas. Also, different values of ϵ_r (ranging from 4 to 100) can be used, thus allowing the designer the flexibility in controlling the size and bandwidth.
- Depending on the resonator shape, various modes can be excited within the DRA element. These modes can produce different radiation patterns for various coverage requirements. Also, the Q -factor of some of these modes will depend on the aspect ratio of the DRA, thus allowing one more degree of flexibility in the design.

- Many of the existing feeding schemes can be used (slots, probes, microstrip, coplanar waveguides, dielectric image guide, etc.). This makes them easy to integrate with existing technologies.
- Compared with the microstrip antenna, DRA has a much wider impedance bandwidth. This is because the microstrip antenna radiates only through two narrow radiation slots, whereas the DRA radiates through the whole antenna surface except the grounded part. Moreover the operating bandwidth of a DRA can be varied by suitably choosing the dielectric constant of the resonator material and its dimensions.
- DRAs have been designed to operate over a wide frequency range (1 GHz to 44 GHz) compared with other antennas existing in the literature.
- DRAs have a high dielectric strength and hence higher power handling capacity. Moreover the temperature-stable ceramics enable the antenna to operate in a wide temperature range.

Fundamental Modes and Their Radiation Mechanisms

A microwave resonator has an infinite number of resonant modes, each corresponding to a particular resonant frequency at which the stored electric energy is equal to the magnetic energy. To simplify the discussion, we use the cylindrical resonator as our reference in the free space. Based on that, the excited modes can be classified into three distinct types: *TE*, *TM*, and hybrid. The fields for *TE* and *TM* modes are axisymmetric, whereas hybrid modes are azimuthally dependent. According to the mode nomenclature described by Kobayashi et al.,^{18,19} the variation of fields along the azimuthal, radial and Z-direction inside the resonator is denoted by adding mode indices as subscripts to each family of modes. The *TE*, *TM*, and hybrid modes are classified as $TE_{mnp+\delta}$, $TM_{mnp+\delta}$ and $HE_{mnp+\delta}$ respectively. The first index denotes the number of full-period field variations in azimuthal direction, the index n ($n = 1, 2, 3, \dots$) denotes the order of variation of the field along the radial direction and the index $p + \delta$ ($p = 0, 1, 2, \dots$) denotes the order of variation of the fields along the Z-direction. Here, the third index denotes the fact that the dielectric resonator is shorter than integer multiples of half the dielectric wavelength. The actual value of δ depends on the relative dielectric constant of the resonator and the substrate and on the proximity to the top and bottom conductor planes. An interesting feature of DR is the variation in field distribution of different modes, because the modes behave like electric and magnetic multipoles such as dipole, quadrupole, octupole, etc. The mode nomenclature makes possible the accurate prediction of far-field radiation of dielectric resonators in their application as antennas. For example, the $TE_{01\delta}$ mode radiates like a short magnetic dipole oriented along its axis, whereas its *TM* counterpart radiates like a short electric dipole and the $HE_{11\delta}$ mode radiates like a horizontal magnetic dipole. The modes that are most commonly used for radiating applications are $TE_{01\delta}$, $TM_{01\delta}$, and $HE_{11\delta}$. The corresponding field distributions for cylindrical DRs are shown in Figure 17-1.

The equations for the resonant frequency and radiation *Q*-factor of the $TE_{01\delta}$, $TM_{01\delta}$ and $HE_{11\delta}$ modes of a cylindrical DRA with dielectric constant ϵ_r , radius a , and height h kept above an infinite ground plane and fed by an electromagnetic signal of free space velocity c are given next. These equations have been derived by a combination of dielectric waveguide model and curve fitting ($k_o a$ vs. a/h) based on extensive numerical and experimental investigations.^{15,20}

$TE_{01\delta}$ mode:

$$f_o = \frac{2.921c\epsilon_r^{-0.465}}{2\pi a} [0.691 + 0.319x - 0.035x^2] \quad (17-1)$$

$$Q = 0.012\epsilon_r^{1.2076} [5.270x + 1106.188x^{0.625} e^{-1.0272x}] \quad (17-2)$$

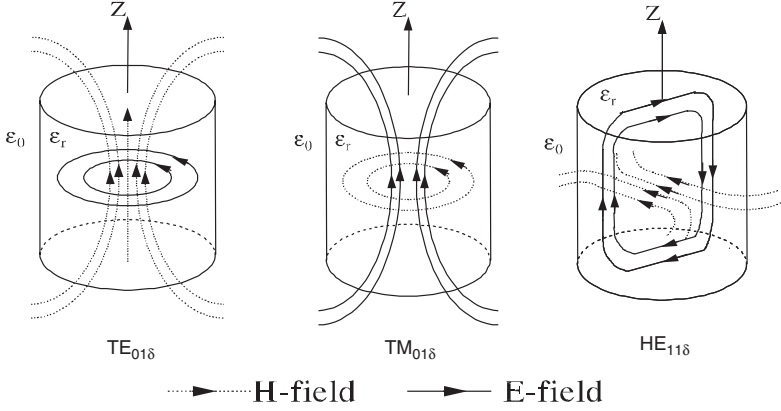


FIGURE 17-1 Sketch for the field distributions of the first three common modes

where $x = a/2h$. The ideal radiation pattern of $TE_{01\delta}$ mode is as shown in Figure 17-2, if the DR is split along its axis and placed on a ground plane.

$TM_{01\delta}$ mode:

$$f_o = \frac{2.933c\epsilon_r^{-0.468}}{2\pi a} \{1 - [0.075 - 0.05x](\epsilon_r - 10)/28\} \{1.048 + 0.377x - 0.071x^2\} \quad (17-3)$$

$$Q = 0.009\epsilon_r^{0.888} e^{0.040\epsilon_r} \{1 - [0.3 - 0.2x](38 - \epsilon_r)/28\} \{9.498x + 2058.33x^{4.322} e^{-3.501x}\} \quad (17-4)$$

The ideal radiation pattern of $TM_{01\delta}$ mode when it is cut to half and placed on a ground plane is depicted in Figure 17-3.

$HE_{11\delta}$ mode:

$$f_o = \frac{2.735c\epsilon_r^{-0.436}}{2\pi a} [0.543 + 0.589x - 0.050x^2] \quad (17-5)$$

$$Q = 0.013\epsilon_r^{1.202} [2.135x + 228.043xe^{-2.046x+0.111x^2}] \quad (17-6)$$

The ideal radiation pattern of $HE_{11\delta}$ mode when it is cut to half and placed on a ground plane is illustrated in Figure 17-4.

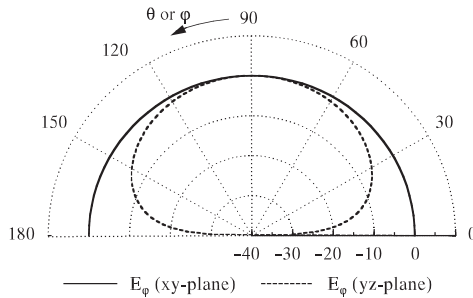


FIGURE 17-2 Ideal radiation pattern of $TE_{01\delta}$ mode

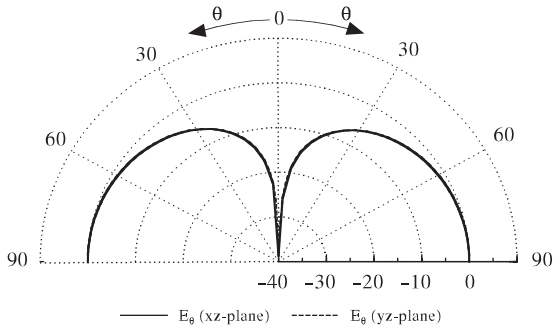


FIGURE 17-3 Ideal radiation pattern of $TM_{01\delta}$ mode

The resonant frequencies and radiation of ideal modes for other geometries can be obtained with some approximations or with complete numerical eigenvalue solutions. The exception are hemispherical shapes for which modes can be obtained exactly by enforcing the boundary conditions on the surface of the sphere to obtain a characteristic (eigenvalue) equation from which numerical solutions can be obtained to get the resonant frequencies.

Excitation Techniques of DRAs

The operational mode of a DRA depends on the method of excitation employed. Moreover, the coupling mechanisms significantly affect the resonant frequency and radiation Q -factor of a DRA. A number of excitation techniques have been adopted. These include coaxial probe,^{21,22} aperture coupling with a microstrip feedline,^{23,24} aperture coupling with a coaxial feedline,^{25,26} waveguide coupled aperture,^{27,28} direct microstrip feedline,^{29,30} coplanar feed,³¹ soldered through probe,³² slot line,³³ conformal strip³⁴ and direct image guide.³⁵ Some of these excitation methods are illustrated in Figure 17-5.

In aperture coupling methods, the aperture can be of any shape, such as narrow slot, loop, cross, or C -shape cut in the ground plane and can be fed by a microstrip line/coaxial feedline beneath the ground plane or aperture cut on the surface of a waveguide. Here the aperture behaves like a magnetic current running parallel to the length of the slot, which excites the magnetic fields in DRA. In this case moving the DRA with respect to the

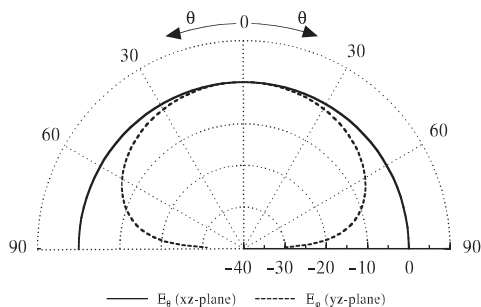


FIGURE 17-4 Ideal radiation pattern of $HE_{11\delta}$ mode

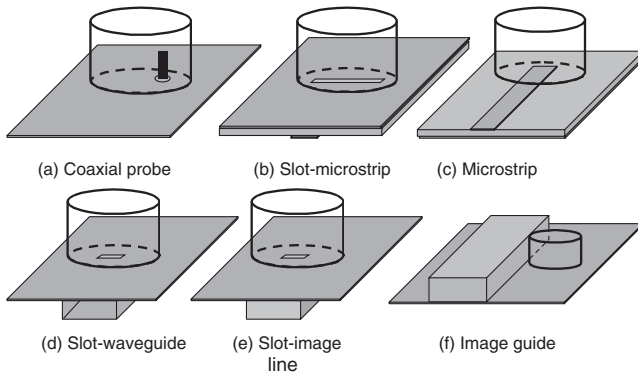


FIGURE 17-5 Different excitation mechanisms for the DRA

aperture or controlling the aperture size can adjust the coupling level. This method has the advantage of feeding networks kept below the ground plane and hence avoids the spurious modes. Moreover, slot aperture coupling is widely used for integrating DRAs with printed feed structures.

The excitation of DRs using coplanar waveguide lines appears to be highly promising because they enable easy coupling with MMICs. In this case, coupling level can be adjusted by positioning the DRA over the coplanar structure. Impedance tuning can be done by adding stubs, slots, or loops at the end of the coplanar line. Though coaxial probes and coplanar loops work alike, the latter has the advantage of unobtrusive and planar structure to benefit integration. By moving the position of DR over the loop, the operational mode can be selected.

Dielectric image guide excitation of DRAs offers advantages over microstrip line methods in that they do not suffer from conductor losses, especially at millimeter wave frequencies. Here the coupling between the guide and the DR is usually small, but can be increased by operating the guide closer to the cutoff frequency. This method is similar to the waveguide slot method, but with the image line replacing the waveguide. These types of excitations find many applications, especially in series fed linear DRA arrays.

In coaxial probe feed excitation, the center pin of a coaxial transmission line is extended through the ground plane. It can also be soldered to a flat metal strip. This can be considered as an electric current running vertical to the DRA height. The probe can either be located adjacent to the DRA or can be embedded within the body. The strength of coupling depends on the length and location of the probe. Various modes can also be excited based on the probe location. The extent of coupling can be adjusted by modifying the probe height. In this method, various modes are excited depending on the position of the probe. Another advantage of this method is that the antenna system can be directly connected to a 50 Ω circuit without the aid of any matching network.

Use of the direct microstrip line mechanism is the simplest method to energize DRAs. In this method, a microstrip line printed on the same substrate excites a DR that could be placed directly over the microstrip line or nearby over the dielectric substrate. The level of coupling can be adjusted by the lateral position of the DR with respect to the microstrip line and by using substrate with different permittivity. For wider bandwidth, ϵ_r of the substrate should be kept low, but requires a reasonable value for better coupling. To have acceptable radiation efficiency, microstrip-line excited DRA arrays are being used. Microstrip line offers the advantage of easy and cost-effective fabrication of DRA arrays because feedlines can be

simply printed over the substrate. But these have the disadvantage that polarization of the array is dictated by the orientation of the microstrip line. Moreover, this excitation scheme may also generate surface waves in the dielectric substrate, which is highly undesirable.

It should be clear that once we use the real source, different modes could be excited and in some cases the radiation patterns could be deformed from the intended mode pattern. Such deformation may be controlled or reduced by adjusting the excitation position or coupling level.

17.2 POSSIBLE METHODS OF ANALYSIS

The approaches to solving DRA problems theoretically can be broadly classified into two: (1) theoretical models like cavity, waveguide, and curve fitting methods and (2) numerical methods based on frequency domain and time domain techniques.

Theoretical Models

Theoretical models are appropriate for conical structures. Then modal analysis and Green's functions can be derived getting analytical expressions.

Cavity Model (Cylindrical DRA) Cavity model¹² is the first and simplest analysis technique applied to a DRA. As an approximation, the DRA can be considered as a dielectric cavity with the surfaces as perfect magnetic conductors. In that case, the *TE* and *TM* modes are the only modes possible. This model is used to compute the resonant frequency and the radiation patterns. Because of this approximation, high error is expected in predicting the resonant frequency.

Dielectric Wave-Guide Model (Rectangular DRA) This method was first employed³⁶⁻³⁸ to analyze the resonant frequency and radiation fields of a rectangular DRA. Consider a rectangular DRA of length l , width w , and height h , mounted on an infinite ground plane. Typically the $TE_{11\delta}^z$ mode will be the prominent one excited, for $l, w > h$. As an approximation the top surface and two sidewalls (mostly in the propagation direction) assume perfect magnetic boundaries, whereas the other two sidewalls assume imperfect magnetic walls and the bottom surface that is in contact with the ground plane acts as an electric wall. With this, the electric and magnetic components of the field in the DRA can be expressed as

$$H_x = \frac{(k_x k_z)}{j\omega\mu_o} \sin(k_x x) \cos(k_y y) \sin(k_z z) \quad (17-7)$$

$$H_y = \frac{(k_y k_z)}{j\omega\mu_o} \cos(k_x x) \sin(k_y y) \sin(k_z z) \quad (17-8)$$

$$H_z = \frac{(k_x^2 + k_y^2)}{j\omega\mu_o} \cos(k_x x) \cos(k_y y) \cos(k_z z) \quad (17-9)$$

$$E_x = k_y \cos(k_x x) \sin(k_y y) \cos(k_z z) \quad (17-10)$$

$$E_y = -k_x \sin(k_x x) \cos(k_y y) \cos(k_z z) \quad (17-11)$$

$$E_z = 0 \quad (17-12)$$

where

$$k_x^2 + k_y^2 + k_z^2 = \epsilon_r k_o^2, \quad k_z \tan(k_z l/2) = \sqrt{(\epsilon - 1)k_o^2 - k_z^2} \quad (17-13)$$

$$k_o = \frac{2\pi f_0}{c}, \quad k_x = \frac{n\pi}{l}, \quad k_y = \frac{m\pi}{m} \quad (17-14)$$

The transcendental equation (see Eq. 17-13) is solved to yield the resonant frequency of the radiating mode. A modified dielectric waveguide model has been proposed³⁹ to yield more accuracy in the calculation of the resonant frequency.

Cylindrical Ring Resonator In the case of a cylindrical ring⁴⁰ DRA of outer diameter $2a$, inner diameter $2b$, and height h , the resonant frequency of the $TM_{01\delta}$ can be given⁴¹ as

$$f_0 = \frac{c}{2\pi\sqrt{\epsilon_r}} \sqrt{\left(\frac{\pi}{2h}\right)^2 + \left(\frac{X_0}{a}\right)^2} \quad (17-15)$$

where X_0 is represented as

$$\frac{J_1(X_0)}{Y_1(X_0)} = \frac{J_1\left(\frac{b}{a}X_0\right)}{Y_1\left(\frac{b}{a}X_0\right)} \quad (17-16)$$

where $J_1(x)$ and $Y_1(x)$ are the first-order Bessel functions of the first and second kind, respectively. By solving Eq. 17-16, one could have a plot of X_0 as a function of b/a . The normalized Q -factor is obtained as a ninth order polynomial equation.

Numerical Methods

Numerical methods can be classified into frequency domain methods (such as spectral domain method, modal expansion, and method of moments) and time domain techniques (like FDTD, TLM, and FEM). Some of these techniques adopted for the analysis of DRAs that appeared in literature are shown next.

- MOM to investigate the effect of air gaps of cylindrical DRAs⁴²
- MOM to calculate internal field patterns of various modes of CRDA⁴³
- Model expansion used to compute mutual impedance between probe-fed hemispherical DRAs⁴⁴ and split CDRAs⁴⁵
- Use of mode matching techniques and Green's functions for hemispherical DRAs^{46,47}
- FDTD used to calculate circularly polarized patterns of cross-shaped DRAs⁴⁸
- FDTD methods to find input impedance of slot-fed RDRA⁴⁹
- TLM methods to compute input impedance of microstrip-fed multisegment DRAs⁵⁰
- FEM-based simulation techniques like HFSS
- Hybrid approach (modal expansion and spectral domain method) to analyze aperture-coupled DRA¹⁷

Green's Function Analysis of Hemispherical DRA The hemispherical DRA can be analyzed by using the Green's function of the spherical structures. Such analysis enables deriving Green's function for multilayer spherical DRA.

17.3 PROGRESS IN DRA RESEARCH

Various Geometries of DRAs

One of the attractive features of a DRA is that it can assume a number of shapes. Moreover the mode of operation and performance of a DRA can be varied by selecting a DR with desired structure. Hence a number of DRA geometries have been tried experimentally. The first systematic, theoretical, and experimental study was made on cylindrical¹² disk DRA geometry. Later geometries such as split cylinder,⁵¹ sectorized cylinder,⁵² cylindrical rings,⁵³ metallized DRAs,⁵⁴ triangular,⁵⁵ rectangular,¹³ notched rectangular DRA,⁵⁶ chamfered DRA,⁵⁷ conical,⁵⁸ elliptical,⁵⁹ spherical,¹⁴ hemispherical,⁶⁰ spherical cap,⁶¹ tetrahedral,⁶² perforated DRA,⁶³ stepped DRAs,^{64–66} and hybrid DRAs,^{67–71} have been reported. Some typical examples are illustrated in Figure 17-6. It was found that⁵³ DRAs operating at their fundamental modes radiate like an electric or magnetic dipole, which depends on the mode of excitation and geometry of the bulk dielectric material. Though several geometries have been introduced, the most studied and common structures are still the cylindrical and rectangular DRAs because of the simplicity in their design, fabrication, and analysis.

Wideband Techniques

The important and most common methods adopted for the fabrication of wideband DRAs can be broadly classified as single element methods, multiple DRAs, hybrid antennas, and DRA arrays. The single element method includes:

- Optimizing the feeding mechanisms and the DRA parameters
- Use of modified feed geometries (stub matching)
- Changing the shape of the DRA

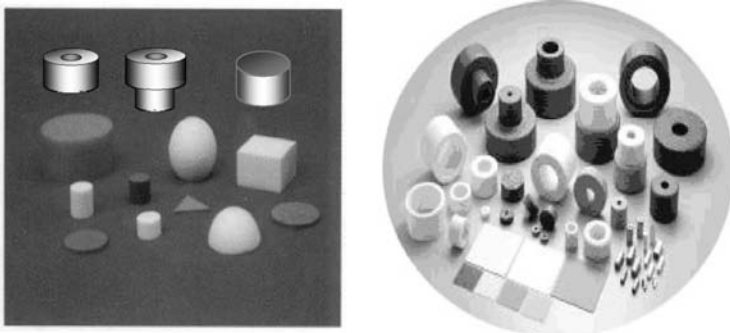


FIGURE 17-6 Pictures of DRs with various shapes

- Introduction of air gap between the ground plane and DR
- Reducing the dielectric constant of DR
- Use of parasitic coupling with different resonators

There are reports⁷²⁻⁷⁴ regarding the bandwidth enhancement of rectangular, cylindrical, and elliptical DRAs employing microstrip matching stub. Though the bandwidth is enhanced, this technique worsens the cross-polarization level, which is undesirable in many applications.

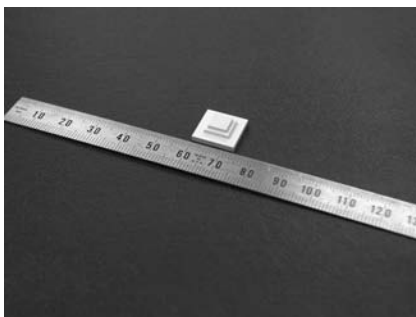
The DRA geometries have been tailored to yield better performance. The different DRA geometries reported are described in the preceding subsection. A wide range of bandwidths has been reported based on the feeding techniques and DRA parameters. Modification of the DRA structure (see Figure 17-7) in the form of a stepped pyramid has been reported to have an ultra wideband as high as 60 percent.⁶⁶ The similar stepped stair structure has also been reported with cylindrical DRA.⁶⁷

Conical⁷⁵ and split conical⁷⁶ DRAs shown next (see Figure 17-8) are also reported as potential wideband candidates. A further enhanced bandwidth was observed with inverted conical DRAs. Split conical DRAs⁷⁷ offer still higher bandwidth (about 50 percent for DR of $\epsilon_r = 12$, $a = 22.5$ mm, $b = 49$ mm, and $h = 52$ mm), as shown in Figure 17-9. However, the split structures have the drawback of exciting higher-order modes, which perturbs uniform radiation within the middle of the band, as shown in Figure 17-10.

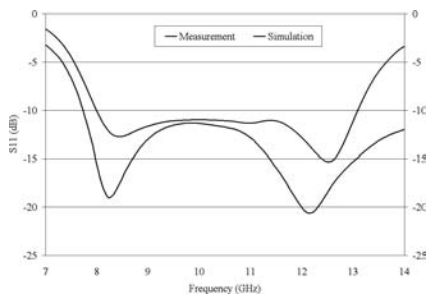
Another novel single-element DRA structure that attracted attention because of higher bandwidth is tetrahedral. An inverted (truncated) tetrahedral DRA⁷⁸ as shown in Figure 17-11 exhibits a wide bandwidth of about 40 percent with parameters as described in Figure 17-12. Compared with conical DRA, this structure has the advantage of a uniform radiation pattern (see Figure 17-13).

Yet another novel geometry (shaped DRA), which has attracted recent attention, is shown in Figure 17-14. The geometry is an inverted L -shaped cylindrical DRA, which is examined⁷⁹ for wireless applications covering four wireless bands as labeled in Figure 17-15. About 40 percent impedance bandwidth with 8 dBi peak gain is demonstrated.

Introduction of air gaps between the ground plane and DRA has proved⁸⁰ an efficient method to enhance the bandwidth. This normally results in an increase of resonant frequency and decrease in Q -factor. This method could attain an impedance bandwidth of up to 40 percent.⁸¹ Insertion of low-permittivity substrates in place of air gaps was also tried.⁸²



(a)



(b)

FIGURE 17-7 (a) Geometry; (b) Measured vs. computed reflection coefficients of stepped DRA

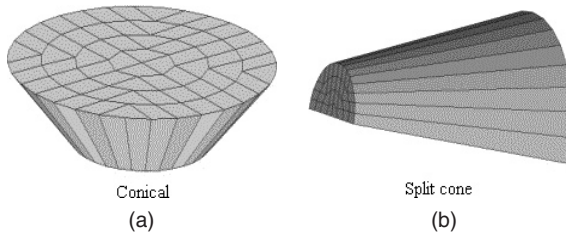


FIGURE 17-8 Conical and split conical DRA geometries for wide-band operation

Use of cylindrical rings⁸³ instead of cylindrical DRAs reduces the effective dielectric constant and hence increases the bandwidth. In this case also, introduction of an air gap between the DRA and the ground plane⁸⁴ can further improve bandwidth.

Significant improvements in bandwidth of rectangular DRAs have been reported by the use of notched structure⁸⁵ (see Figure 17-16). The idea is to effectively lower the dielectric constant and Q -factor to enhance bandwidth. Further tuning in dimensions can result in a dual band operation. With this technique a notched rectangular DRA with 32 percent bandwidth is reported, while the solid equivalent structure (rectangular DRA) resulted in only 10 percent bandwidth. In addition to the increased bandwidth, this geometry needs a hole drilled to insert the probe, which results in easier fabrication.

The second class of broadband DRA fabrication (multiple DRAs) methods consists of

- Loaded DRA with different dielectric materials
- Cavity backed DRAs
- Stacked DRAs
- Embedded DRAs
- Hybrid antennas (include DRA loaded with microstrip patch, DRA loaded monopole, etc.)

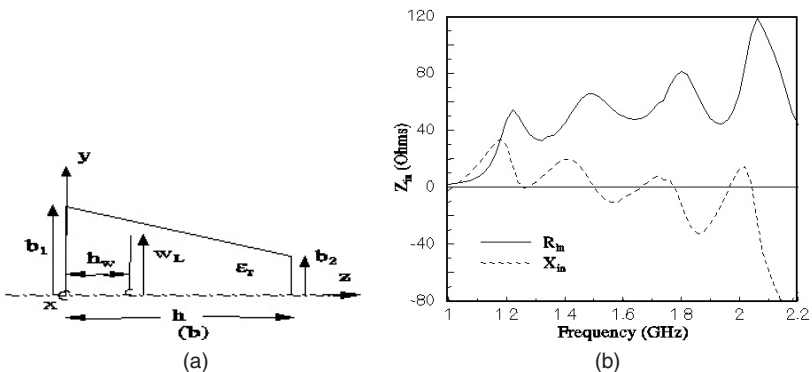


FIGURE 17-9 Split cone DRA (a) parameters and (b) its input impedance

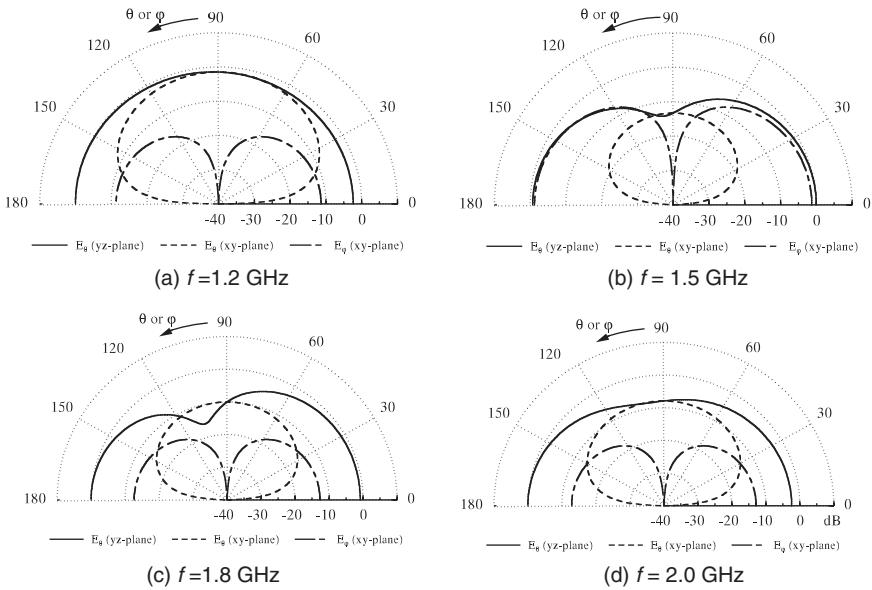


FIGURE 17-10 Radiation patterns of the antenna in Figure 17-9, with $b_1 = 4.9$ cm, $b_2 = 2.25$ cm, $h = 5.2$ cm, $w_L = 1.92$ cm, $h_w = 0.5$ cm, and $\epsilon_r = 12$ at different frequencies

Disk loaded DRAs⁸⁶ were reported mainly for frequency tuning and bandwidth improvement. The loaded disk can be either metallic or dielectric. However, in either case the disk perturbs the resonant frequency of the DRA; hence, disk size should be carefully selected.

Cylindrical DRAs backed by hemispherical cavity have been used⁸⁷ as a broadband antenna. The structure yielded 38 percent impedance bandwidth. The only drawback of this antenna is the bulky structure.

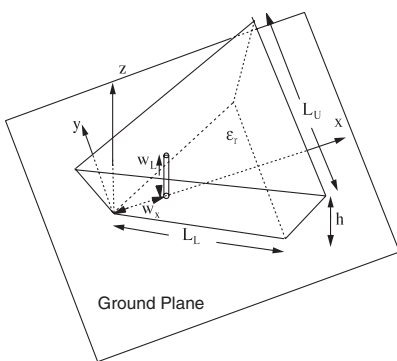


FIGURE 17-11 Geometry of truncated tetrahedral DRA

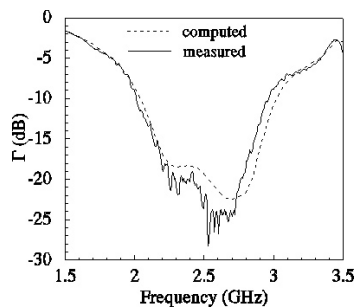


FIGURE 17-12 Computed and measured reflection coefficients of a truncated tetrahedron with $L_U = 6.4$ cm, $L_L = 2.5$ cm, $h = 2.6$ cm, $w_x = 0.55$ cm, $w_L = 1.15$ cm, and $\epsilon_r = 12$

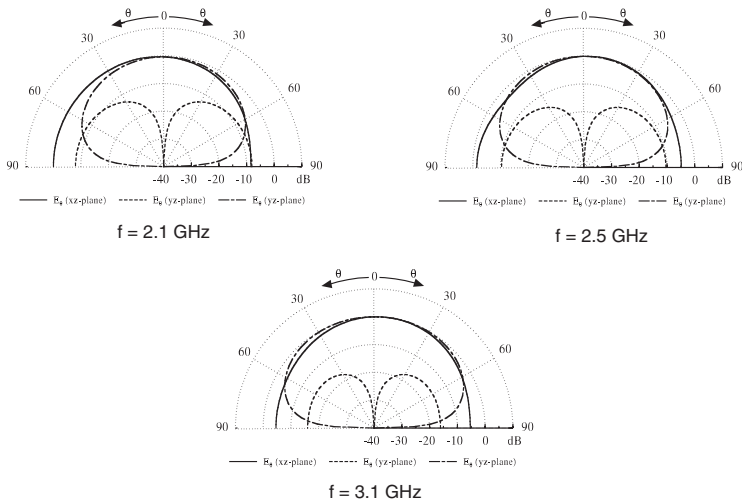


FIGURE 17-13 Computed radiation patterns of a truncated tetrahedron at different frequencies with $L_U = 6.4$ cm, $L_L = 2.5$ cm, $h = 2.6$ cm, $w_x = 0.55$ cm, $w_L = 1.27$ cm, and $\epsilon_r = 12$

Stacking of DRAs^{88–90} was reported as an efficient technique to improve the gain, bandwidth, and radiation performances. Further increase in bandwidth can be achieved by the introduction of air gaps. Another advantage is that a single probe can be used for the excitation of all DRs. Bandwidth in the range of 25–35 percent has been reported with this geometry.

Embedding of one DRA in another is a new method for bandwidth enhancement. This can be done either with both the DRs having the same permittivity,⁹¹ or with the inner and outer DRAs having different permittivity.⁹² The latter has reported an impedance bandwidth of over 50 percent, when $a_2 = 4$ mm, $h_2 = 2$ mm, and $\epsilon_r = 12.3$ and $a_1 = 5.5$ mm,



FIGURE 17-14 Shaped DRA: DCS-WLAN bands

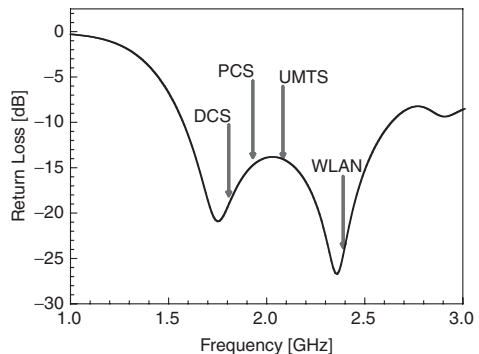


FIGURE 17-15 Return loss characteristics of shaped DRA with possible bands of operation

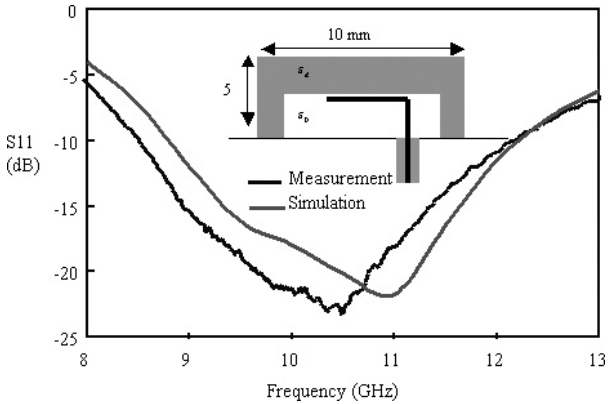


FIGURE 17-16 Geometry and measured return loss of notched DRAs

$h_1 = 6$ mm, and $\epsilon_{r1} = 4.1$. The antenna is placed on top of a narrow slot with length 5.5 mm and width 0.5 mm. The slot is excited parasitically with a $50\ \Omega$ microstrip line on a substrate of thickness 0.813 mm, dielectric constant 3.38, and extended stub of 4 mm.

It has also been observed that the geometry with higher dielectric constant material touching the slot provides better matching.

Mainly two types of DRA-based hybrid antenna structures are reported: DRA-loaded microstrip patch antennas and DRA-loaded monopole antennas.

Microstrip patch antennas possess many advantages like light weight, compactness, high gain, low profile, easy fabrication, etc. However, they suffer from conduction losses and low-impedance bandwidth. There are reports that the bandwidth performance of a microstrip antenna can be improved by loading with a DRA.^{71,93} Bandwidth of about 24 percent has been achieved with this technique. Patch antennas loaded with stacked DRAs⁹⁴ were also reported to yield better bandwidth.

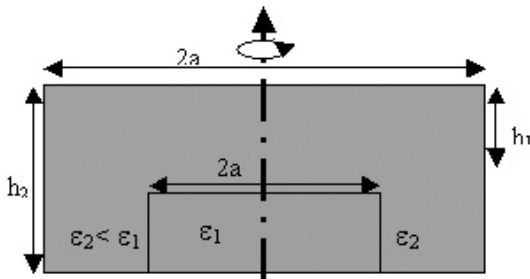


FIGURE 17-17 Embedded DRA

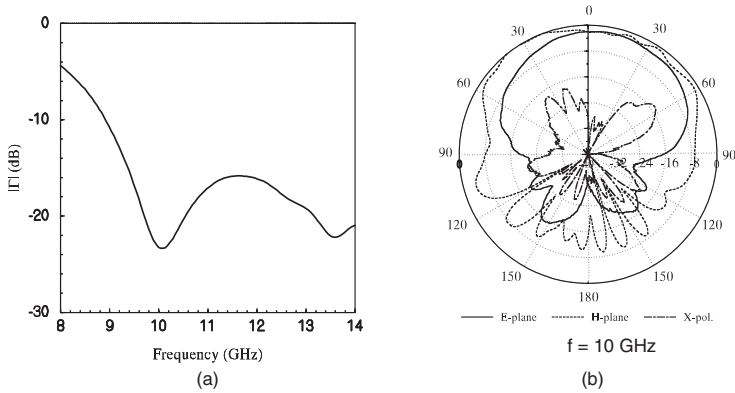


FIGURE 17-18 (a) Measured reflection coefficients and (b) radiation patterns of embedded DRA excited by narrow slot

One of the recent techniques developed to have an ultra wideband hybrid antenna⁶⁹ is shown in Figure 17-19. Here a dielectric ring resonator surrounds a monopole antenna. Both are centered at the same axis and the monopole acts as a quarter-wavelength radiator as well as a feed for DRA. The DRA is designed to excite in the dominant $TM_{01\delta}$ mode, which has a circular symmetric field distribution similar to that of a monopole. A further in-depth investigation on the design parameters of this structure was also reported.⁹⁵ For wideband operation the monopole is designed to operate at the lower end of the frequency spectrum, while DRA operates toward the upper end. The design parameters can be selected to have separate frequencies, so that dual band operation occurs.

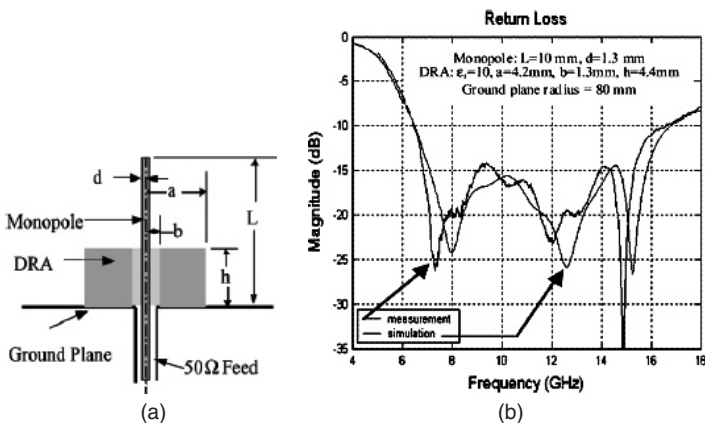


FIGURE 17-19 (a) Cross sectional view of monopole-fed DRA and (b) wideband characteristics

More recently, a compact coplanar waveguide-fed planar monopole-DRA structure has been reported⁹⁶ with about 50 percent bandwidth when a DR of permittivity 38 was used (see Figure 17-20).

17.4 DRA ARRAYS

DRA elements of proper geometry can be assembled and fed in a suitable way to modify their gain, bandwidth and radiation performances and hence can be advantageous over a single element. These DRA arrays^{97–102} find applications in terrestrial applications as well as radars. The performance of an array depends on the geometry and dimensions of individual elements, number of elements, their spacing, mode of operation and feeding techniques adopted. The developed DRA arrays can be broadly classified as linear DRA arrays and planar arrays.

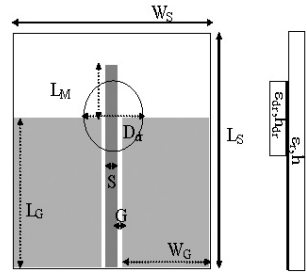


FIGURE 17-20 Geometry of hybrid CPW-fed planar monopole/dielectric resonator antenna. $L_G = 25$ mm, $W_G = 10$ mm, $L_M = 18$ mm, $\epsilon_r = 4.4$, $h = 1.6$ mm, $L_S = 50$ mm, $W_S = 30$ mm, $S = 3$ mm, $G = 0.3$ mm, $\epsilon_{dr} = 38$, $h_{dr} = 3.83$ mm, $D_{dr} = 16.4$ mm

Linear DRA Arrays

A linear DRA array produces radiation patterns with narrow beamwidth in a plane parallel to the array axis and a broad beamwidth in the orthogonal plane. The phase and amplitude of excited radiation in the element can be controlled to have a specific beam shape. The most common excitation technique employed in this case is microstrip line methods. Linear DRA array excitation can be of (1) microstrip parallel feed, (2) microstrip series feed, (3) microstrip branch line feed, and (4) dielectric image guide feed. Figure 17-21 shows¹⁰³ the radiation patterns of a uniformly excited 32-element linear array of dielectric resonator disks. The elements are separated by a half wavelength, excited by coaxial probes, and arranged in the E -plane of the array. The radiation patterns are shown both without mutual coupling effects and with the mutual coupling effect. The radiation patterns are scanned at 0° , 30° , and 60° . It can be seen that

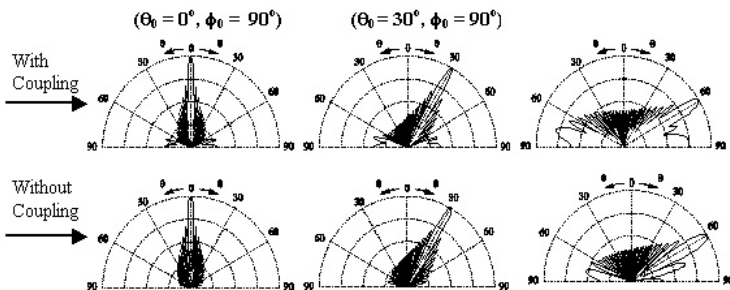


FIGURE 17-21 E -plane radiation patterns of 32 DRA elements with and without mutual coupling effect

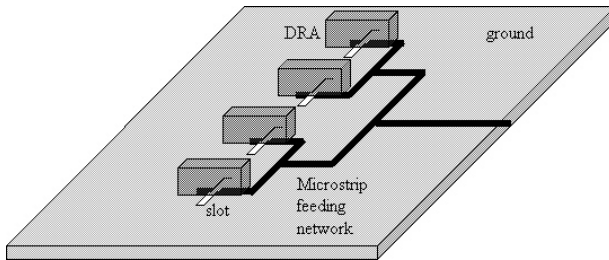


FIGURE 17-22 Linear arrays of DRAs excited with corporate feed network

the desired scanning direction is achieved. The mutual coupling shows its effect with the low elevation angles from 60° to 90° .

Microstrip Parallel Feedline

This type of feed results in a more compact and smaller antenna network loss. Adjusting the spacing between the DRA and the stripline can control the coupling between them. Consider a linear array¹⁰⁴ as shown in Figure 17-22.

In this case four rectangular DRAs are fed via a slot aperture using a microstrip parallel network. The antenna system operates at 40 GHz. The measured gain is 6.2 dBi. The main disadvantage of this kind of feeding is the conductor and surface wave losses from the microstrip line.

Microstrip Series Feedline The use of series feeding results in a more compact and lower-loss antenna system compared with parallel feeds. However, series feeding suffers from reduced bandwidth due to increased phase errors from the designed frequency, since electrical path lengths are a function of frequency. In this case, two approaches can be applied for designing: the resonant approach¹⁰⁵ and the leaky-wave approach.¹⁰⁶ In the resonant approach (see Figure 17-23), the microstrip line terminates to an open circuit, which creates standing waves on the line.

In this case, when the number of elements is small, the amount of coupling from the microstrip line to each DRA element is low and this method better predicts radiation patterns. On the other hand, if the number of elements is high, very little energy will be left

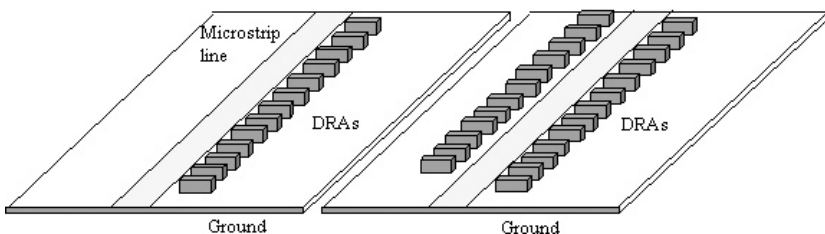


FIGURE 17-23 Linear array of DRAs with series fed microstrip lines

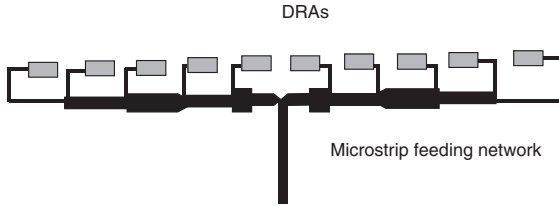


FIGURE 17-24 Linear DRA array fed by branched microstrip lines

over to reach the open termination point to create the standing waves. In such cases, the radiation patterns can be more accurately predicted using leaky wave analysis.

The main drawback of series and parallel fed microstrip arrays is the need for more elements for better radiation performance, making the antenna system bulky.

Microstrip Branch Line Feed Branched microstrip line feed has been proposed as an alternative for series and parallel fed antenna arrays, which results in better coupling of energy between the elements and transmission line. An example of this type of DRA array¹⁰⁷ is depicted in Figure 17-24. Furthermore, this method enables a much more precise amplitude control than previous methods.

Dielectric Image Guide Feed The use of a low-loss dielectric image guide as a transmission line can reduce the losses due to conductor and surface waves that are generated in microstrip lines. This method could produce a significant radiation performance (especially in gain and efficiency) of a DRA array. A linear DRA array fed by a dielectric image guide¹⁰⁸ is shown in Figure 17-25.

However, it was observed that this kind of feeding could still result in lesser gain compared with a microstrip line unless transition was carefully designed. This is because

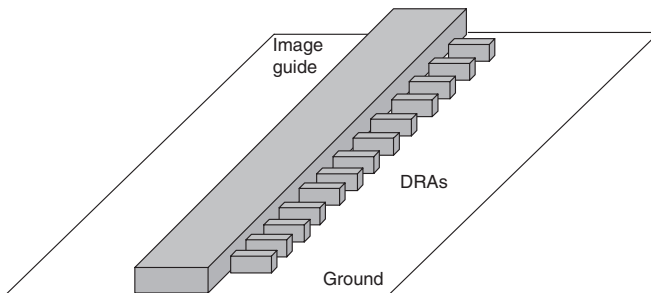


FIGURE 17-25 Dielectric image guide fed DRA array

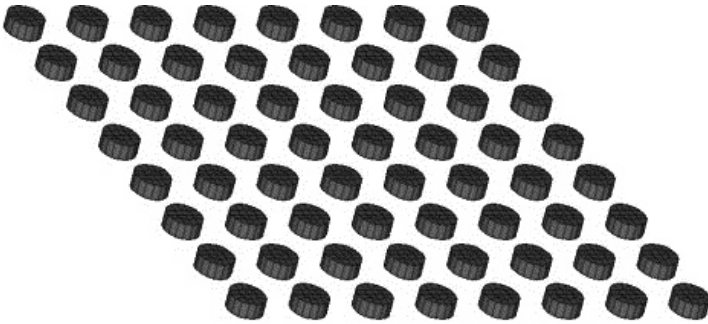


FIGURE 17-26 Geometry of 8×8 planar DRA array

the spurious radiation from the feed transition into the dielectric image guide could degrade the benefits of low-loss dielectric.

Planar DRA Arrays

An example of an 8×8 array of DRA is provided to show the effect of the mutual coupling. The array geometry is shown in Figure 17-26. The radiation patterns for the array are shown in Figure 17-27. Again the effect of the mutual coupling is noticed only in the radiation patterns at the low elevation angle below 60° .

Figure 17-28 shows the geometry and the radiation patterns of a 4×4 stacked DRA array. Narrow slots excite the array. The radiation patterns show a front-to-back ratio of 16 dB.

One of the practical problems of the DRA array is assembling the individual elements and aligning them. Perforation techniques can be used to fabricate an array of DRAs from a single dielectric sheet, as shown in Figure 17-29. This also eliminates the difficulty of placing and bonding individual elements. Investigations on grid element DRA¹⁰⁹ arrays have shown that this array has better gain and bandwidth performance compared with microstrip fed arrays.

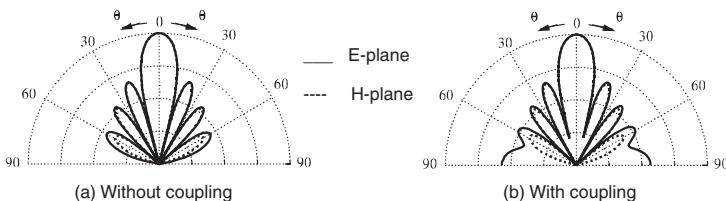


FIGURE 17-27 Radiation patterns of 8×8 uniformly excited planar DRA array

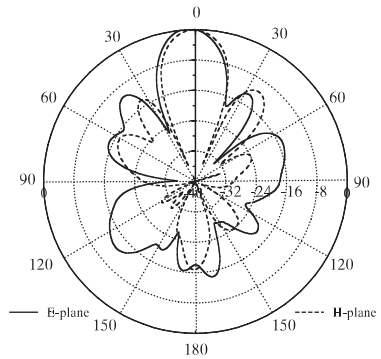
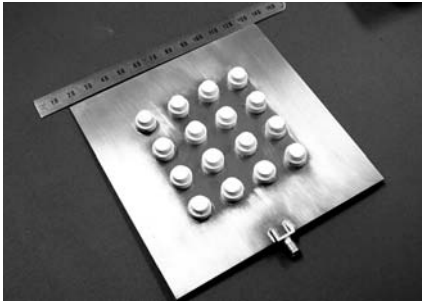


FIGURE 17-28 Geometry and radiation patterns of 4×4 planar stacked DRA array

Sequential Rotation Arrays Sequential rotation arrays¹¹⁰ are mainly intended to improve the axial ratio performance of the DRA elements. Figure 17-30 shows some typical examples of these arrays. The main challenge in the design and fabrication of this type of array is to allocate the feed network that can provide the required amplitude and phase to each radiating element.

Reflect Array DRAs DRA is used as a reflect array operating at 30 GHz. A 529-element reflect array DRA¹¹¹ was machined from a single sheet of dielectric substrate. Later the surface was fused with rigid metallic flat plate. The measured radiation patterns showed again over 24 dB at 30 GHz and sidelobe level below 18 dB.

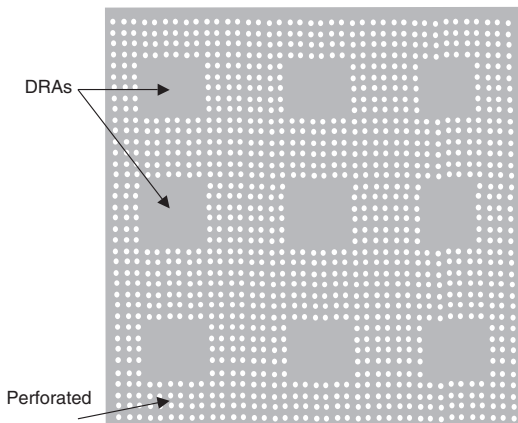


FIGURE 17-29 Perforated DRA array

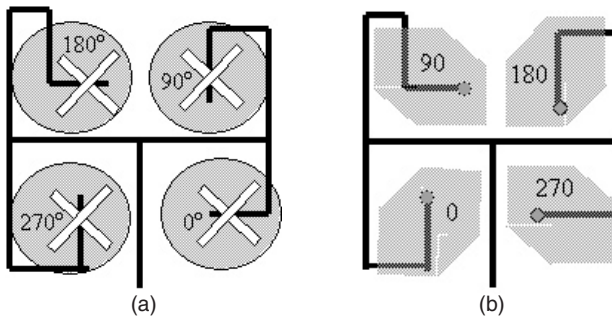


FIGURE 17-30 Sequentially rotated DRA arrays

REFERENCES

1. R. D. Ritchmyer, "Dielectric Resonators," *J. Appl. Phys.*, vol. 10 (1939): 391.
2. A. Okaya, *Proc. IRE*, vol. 48 (1960): 1921.
3. A. Okaya and L. F. Barash, *Proc. IRE*, vol. 50 (1962): 2081.
4. S. B. Cohn, *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16 (1968): 218.
5. D. J. Masse, R. A. Purcel, D. W. Ready, E. A. Maguire, and C. D. Hartwig, *Proc. IEEE*, vol. 59 (1971): 1628.
6. J. Van Bladel, *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23 (1975): 199.
7. W. Wersing, *Electronic Ceramics*, B. C. H. Steele (ed.) (Amsterdam: Elsevier Publishing Co., 1991): 67.
8. D. Kajfez and P. Guillon, *Dielectric Resonators* (Atlanta: Noble Publishing Corp., 1998).
9. J. K. Plourde and C. L. Ren, *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29 (1981): 754.
10. M. Gastine, L. Courtois, and J. L. Dormann, *IEEE Trans. Microwave Theory Tech.*, vol. MTT-15 (1967): 694.
11. O. Sager and F. Tisi, *Proc. IEEE* (September 1968): 1593.
12. S. A. Long, M. McAllister, and L. C. Shen, *IEEE Trans. Antennas Propagat.*, vol. AP-31 (1983): 406.
13. M. McAllister, S. A. Long, and G. L. Conway, *Electron. Lett.*, vol. 19 (1983): 219.
14. M. McAllister and S. A. Long, *Electron. Lett.*, vol. 24 (1988): 1156.
15. R. K. Mongia and P. Bhartia, *International Jour. Microwave and Millimeter-Wave Computer-Aided Engineering*, vol. 4 (1994): 230.
16. A. Petosa, D. J. Roscoe, A. Ittipiboon, and M. Cuhaci, *IEEE Antennas Propagat. Mag.*, vol. 7 (1996): 7.
17. K. M. Luk and K. W. Leung, *Dielectric Resonator Antennas*, J.R. James (ed.) (Baldock, Hertfordshire, UK: Research Studies Press Ltd., 2003).
18. Y. Kobayashi and T. Tanaka, *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (1980): 1077.
19. Y. Kobayashi and M. Miura, *IEEE MTT-S Int. Microwave Symp. Dig.* (1984): 184.
20. A. A. Kishk, A. W. Glisson, and J.P. Junker, *Proc. of the 1999 Antenna Applications Symp.* (Sept 1999): 45-68.
21. S. M. Shum and K. M. Luk, *IEEE Trans. Antennas Propagat.*, vol. 46 (1998): 325.
22. G. P. Junker, A. A. Kishk, and A. W. Glisson, *IEEE Trans. Antennas Propagat.*, vol. 42 (1994): 960.

23. J. T. H. St. Martin, Y. M. M. Antar, A. A. Kishk, and A. Ittipiboon, *Electron. Lett.*, vol. 26 (1990): 2015.
24. K. W. Leung, *IEEE Trans. Antennas Propagat.*, vol. 48 (2000): 1005.
25. K. Y. Chow and K. W. Leung, *IEEE Trans. Electromagnetic Compatibility*, vol. 42 (2000): 290.
26. K. Y. Chow, K. W. Leung, K. M. Luk, and E. K. N. Yung, *IEEE Trans. Antennas Propagat.*, vol. 49 (2001): 307.
27. K. W. Leung and K. K. So, *IEEE Trans. Antennas Propagat.*, vol. 51, no. 9 (2003): 2477.
28. I. Eshrah et al, *IEEE Trans. Antennas Propagat.*, vol. 53, no. 1 (2005): 483.
29. R. A. Kranenburg and S. A. Long, *Electron. Lett.*, vol. 24 (1988): 1156.
30. K. W. Leung, K. Y. Chow, K. M. Luk, and E. K. N. Yung, *Electron. Lett.*, vol. 33 (1997): 1004.
31. R. A. Kranenburg, S. A. Long, and J. T. Williams, *IEEE Trans. Antennas Propagat.*, vol. 39 (1991): 119.
32. K. M. Leung, K. Y. Chow, K. M. Luk, and E. K. N. Yung, *Electron. Lett.*, vol. 33 (1997): 349.
33. H. Y. Lo, K. W. Leung, and K. M. Luk, *Microwave Opt. Technol. Lett.*, vol. 29 (2001): 230.
34. H. Y. Lo and K. W. Leung, *Microwave Opt. Technol. Lett.*, vol. 29 (2001): 317.
35. M. T. Birand and R. V. Gelsthorpe, *Electron. Lett.*, vol. 17 (1981): 633.
36. R. K. Mongia, *IEE Proc.-H*, vol. 139 (1992): 98.
37. R. K. Mongia and A. Ittipiboon, *IEEE Trans. Antennas Propagat.*, vol. 45 (1997): 1348.
38. Y. M. M. Antar and Z. Fan, *IEE Proc. Microwave Antennas Propagat.*, vol. 143 (1996): 113.
39. Y. M. M. Antar et al, *Microwave Opt. Technol. Lett.*, vol. 19, no. 2 (1998): 158.
40. M. Verplanken and J. Van Bladel, *IEEE Trans. Microwave Theory Tech.*, vol. 24, no. 2 (1976): 108.
41. R. K. Mongia et al, *IEE Electron. Lett.*, vol. 29, no. 17 (1993): 1530.
42. D. Kajfez and P. Guillon, *Dielectric Resonators* (Dedham, MA: Artech House, 1986).
43. G. P. Junker, A. A. Kishk, A. W. Glisson, and D. Kajfez, *IEEE Antennas and Propagat. Mag.*, vol. 37 (1995): 40.
44. A. A. Kishk, A. Ittipiboon, Y. Antar, and M. Cuhaci, *IEEE Trans. Antennas Propagat.*, vol. 43 (1995): 198.
45. G. P. Junker, A. A. Kishk, and A. W. Glisson, *IEE Electron. Lett.*, vol. 32, no. 7 (1996): 617.
46. K. W. Leung, K. M. Luk, K. Y. A. Lai, and D. Lin, *Electron. Lett.*, vol. 27 (1991): 2259.
47. K. W. Leung, *IEEE Trans. Antennas Propagat.*, vol. 48, no. 6 (2000): 961.
48. S. M. Shum and K. M. Luk, *IEE Electron. Lett.*, vol. 30, no. 21 (1994): 1727.
49. K. P. Esselle, *URSI Radio Science Meeting*, Newport Beach, CA (June 1995): 28.
50. N. R. S. Simons et al, *Applied Computational Electromagnetic Symposium (ACES-97)*, Monterey, CA, 1997.
51. R. K. Mongia et al, *Microwave and Guided Wave Letters*, vol. 3, no. 2 (1993): 38.
52. M. T. K. Tam and R.D. Murch, *IEEE Antennas Propagat. Symp. Digest*, AP-S (1998): 1958–1961.
53. R.K. Mongia, A. Ittipiboon, P. Bhartia, and M. Cuhaci, *Electron. Lett.*, vol. 29 (1993): 1530.
54. Z. Li, C. Wu, and J. Litva, *IEE Electron. Lett.*, vol. 32, no. 7 (1996): 947.
55. A. Ittipiboon, R. K. Mongia, Y. M. M. Antar, P. Bhartia, and M. Cuhaci, *Electron. Lett.*, vol. 29, (1993): 2001.
56. A. Ittipiboon et al, *IEEE Antennas Propagat. Symp. Digest*, AP-S 1996: 2038.
57. M. Haneishi and H. Takazawa, *IEE Electron. Lett.*, vol. 21, no. 10 (1985): 437.
58. A. A. Kishk, Y. Yin, and A. W. Glisson, *IEEE Trans. Antennas Propagat.*, vol. AP-50 (2002): 469.
59. A. A. Kishk, *Microwave Opt. Technol. Lett.*, vol. 37, no. 6 (2003): 454.
60. K. L. Wong, N. C. Chen, and H. T. Chen, *IEEE Microwave and Guided Wave Letters*, vol. 3, no. 9 (1993): 355.
61. K.W. Leung, K.M. Luk, and E.K.N. Yung, *Electron. Lett.*, vol. 30 (1994): 1366.
62. A.A. Kishk, *IEEE Trans. Antennas Propagat.*, vol. 51, no. 10 (2003): 2913.

63. A. Petosa, A. Ittipiboon, and S. Thirakoune, *IEE Electron. Lett.*, vol. 38, no. 24 (2002): 1493.
64. K. Pliakostathis and D. Mirshekar-Syahkal, *IEEE Antennas Propagat. Symp. Digest*, vol. 2 (2004): 1367.
65. R. Chair et al, *IEEE Antennas Propagat. Symp. Digest*, vol. 2 (2004): 1375.
66. R. Chair et al, *IEE Electron. Lett.*, vol. 40, no. 10 (2004): 581.
67. C. Nannini et al, *Microwave and Opt. Technol. Lett.*, vol. 32, no. 6 (2002): 418.
68. A. Ittipiboon, A. Petosa, and S. Thirakoune, *Symposium on Antenna Technology and Applied Electromagnetics*, ANTEM 2002, (2002): 387.
69. M. Lapierre et al, *IEEE Antennas Propagat. Symp. Digest*, vol 3, AP-S (2003): 16–19.
70. J. George et al, *Microwave Opt. Technol. Lett.*, vol. 17, no. 3 (1998): 205.
71. E. K. N. Yung, W. S. S. Lee, and K. M. Luk, *Microwave Opt. Technol. Lett.*, vol. 7, no. 2 (1994): 55.
72. P. V. Bijumon, S. K. Menon, B. Lethakumari, M. T. Sebastian, and P. Mohanan, *Microwave Opt. Technol. Lett.*, vol. 47, no. 3 (2005): 226–228.
73. P. V. Bijumon, S. K. Menon, M. N. Suma, M. T. Sebastian, and P. Mohanan, *IEE Electron. Lett.*, vol. 41, no. 7 (2005), 385.
74. P. V. Bijumon et al, *Microwave Opt. Technol. Lett.*, vol. 48, no. 1 (2006): 65.
75. A. A. Kishk, A. W. Glisson, and J. P. Junker, *Journal of Electromagnetic Waves and Applications*, vol. 15, no. 6 (2001): 777.
76. A. A. Kishk, A. W. Glisson, and Y. Yin, *Microwave Opt. Technol. Lett.*, vol. 29, no. 3 (2001): 160.
77. A. A. Kishk, Y. Yin, and A. W. Glisson, *IEEE Trans. Antennas Propagat.*, vol. 50, no. 4 (2002): 469.
78. A. A. Kishk, *IEEE Trans. Antennas Propagat.*, vol. 51, no. 10 (2003): 2913.
79. L. C. Y. Chu, D. Guha, and Y. M. M. Antar, *Int. Sym. Antenna Technology and Applied Electromagnetics & Canadian Radio Sciences (URSI/CNC)*, Montreal, July 2006.
80. G. P. Junker, A. A. Kishk, A. W. Glisson, and D. Kajfez, *Electron. Lett.*, vol. 30, no. 3 (1994): 177–178.
81. M. Cooper, M.Sc. thesis, Carleton University, 1(997).
82. A. Laisne, R. Gillard, and G. Piton, *IEE Electron. Lett.*, vol. 37, no. 25 (2001): 1497.
83. R. K. Mongia et al, *IEE Electron. Lett.*, vol. 29, no. 17 (1993): 1530.
84. S. M. Shum and K. M. Luk, *IEE Electron. Lett.*, vol. 30 (1994): 277.
85. A. Ittipiboon et al, *IEEE Antennas & Propagat. Symp. Digest*, AP-S (1996): 2038.
86. Z. Li, C. Wu, and J. Litva, *IEE Electron. Lett.*, vol. 32, no. 7 (1996): 606.
87. K. W. Leung and C. K. Leung, *IEE Electron. Lett.*, vol. 39, no. 14 (2003): 1033.
88. A. A. Kishk, A. Ahn, and D. Kajfez, *IEE Electron. Lett.*, vol. 25, no. 18 (1989): 1232.
89. K. M. Luk, K. W. Leung, and K. W. Chow, *Microwave Opt. Technol. Lett.*, vol. 14, no. 4 (1997): 215.
90. S. M. Shum and K. M. Luk, *IEEE Trans. Antennas Propagat.*, vol. 51, no. 8 (2003): 1996.
91. A. Sangiovanni, J.Y. Dauvignac, and C. Pichot, *IEE Electron. Lett.*, vol. 33, no. 25 (1997): 2090.
92. A. A. Kishk, *Antennas and Wireless Propagat. Lett.*, vol. 4 (2005): 79.
93. P. V. Bijumon et al, *Microwave Opt. Technol. Lett.*, vol. 35, no. 4 (2002): 327.
94. V. Gupta et al, *Microwave Opt. Technol. Lett.*, vol. 37, no. 4 (2003): 302.
95. D. Guha, Y. M. M. Antar, A. Ittipiboon, A. Petosa, and D. Lee, *IEEE Microwave and Wireless Component Lett.*, vol. 5 (2006): 373.
96. M. N. Suma, P. V. Bijumon, M. T. Sebastian, and P. Mohanan, *Proc. of International Conf. on Microwave Materials and Applications*, (June 12–15, 2006).
97. A. Petosa, R. K. Mongia, A. Ittipiboon, and J. S. Wight, *Proc. of IEEE AP-S Conf.* (1995): 1982.
98. Y. X. Guo, K. M. Luk, and K. W. Leung, *IEEE Trans. Microwave Theory Tech.*, vol. 47 (1999): 2146.
99. G. D. Loos and Y. M. M. Antar, *Microwave Opt. Technol. Lett.*, vol. 7, no. 14 (1994): 642.

100. G. Drossos, Z. Wu, and L. E. Davis, *Microwave Opt. Technol. Lett.*, vol. 18, no. 5 (1998): 315.
101. G. P. Junker, A. W. Glisson, and A. A. Kishk, *IEEE APS Intl. Symp. Digest* (1995): 1998.
102. D. Guha and Y. M. M. Antar, *IEEE Trans Antennas Propagat.*, vol. 54, no. 9 (2006): 2657.
103. A. A. Kishk, *Proc. 2003 IEEE Radar Conf.*, (May 5–8, 2003): 268–264.
104. M. G. Keller et al, *Microwave Opt. Technol. Lett.*, vol. 17, no. 6 (1998): 345.
105. A. Petosa et al, *IEE Electron. Lett.*, vol. 32, no. 7 (1996): 608.
106. F. Zucker, “Surface-Wave and Leaky Wave Antennas,” Chap. 16, *Antenna Engineering Handbook*, H. Jasik (ed.) (New York: McGraw-Hill, 1961).
107. A. Petosa et al, *IEE Proc. Microwave Antennas Propagat.*, vol. 144, no. 6 (1997): 472.
108. M. T. Birand and R. V. Gelsthorpe, *IEE Electron. Lett.*, vol. 17, no. 18 (1981): 633.
109. A. Petosa et al, *IEEE Antennas Propagat. Symp. Digest*, vol. 2, AP-S (2005): 168.
110. A. Petosa, A. Ittipiboon, and M. Cuhaci, *IEE Electron. Lett.*, vol. 32, no. 19 (1996): 1742.
111. M. G. Keller et al, *European Microwave Conf.*, (2000): 272.

Chapter 18

Lens Antennas

Donald G. Bodnar

MI Technologies

CONTENTS

18.1	BASIC LENS OPERATION	18-2
18.2	LENS-SURFACE FORMULAS FOR $n > 1$	18-3
18.3	FACTORS AFFECTING GAIN AND SIDELobe LEVELS OF LENS ANTENNAS	18-7
18.4	LENS-SURFACE FORMULAS FOR $n > 1$	18-9
18.5	METAL-PLATE AND WAVEGUIDE MEDIA.	18-11
18.6	WAVEGUIDE LENSES	18-15
18.7	BOOTLACE-TYPE LENSES	18-17
18.8	PHASED-ARRAY LENSES	18-21
18.9	DOmE ANTENNAS.	18-21
18.10	LUNEBURG LENSES	18-23
18.11	FRESNEL ZONE PLATE LENSES.	18-25

18.1 BASIC LENS OPERATION

Lenses and reflectors (see Chapter 15) are used as collimating elements in microwave antennas. Reflectors have a single degree of freedom (the reflector surface), have no internal loss, have low reflection loss, have no chromatic aberration, are relatively easy to support, and can be perforated to reduce weight and wind loads. Lenses have up to four degrees of freedom [inner surface, outer surface, index of refraction n , and (for constrained lenses) inner- versus outer-surface radiator positions], have no aperture blockage by the feed, have internal and surface-reflection losses, must be edge-supported, and are relatively heavy and bulky. In general, if a reflector can provide the required performance for a given application, it should be used. However, since lenses are more versatile, especially in wide-scan-angle performance, they are used in many applications that do not require greater flexibility of a phased array.

The methods for determining radiation-pattern characteristics from an aperture (amplitude and phase) distribution, including deviations from desired distributions, are derived and discussed in Chapter 3. Lens applications are discussed in several other chapters and are only suggested in this chapter. The literature¹⁻⁶ contains summaries of most lens types and list a number of references. The material in Jasik¹ is the basis of this chapter, and many sections are only updated; however, the section on artificial dielectric materials is not repeated here.

Some microwave lenses are adapted directly from optics, but microwave antennas generally use only a single element since multielement designs common in optics are generally too bulky and heavy for microwave applications. However, several techniques are available at microwave frequencies that permit specialized designs which are difficult, if not impossible, to use at optical frequencies; these techniques include nonspherical lens surfaces, artificial dielectric materials, constrained and geodesic media, and variable n with position in the media.

Lenses are designed to collimate one wavefront into another by using ray tracing on the basis of the law of the optical path that all rays between wavefronts (or phase fronts) have equal optical path lengths and on the application of the Fresnel equations (Snell's law plus polarization effects) at the lens surfaces. The lenses discussed in this chapter collimate a spherical or cylindrical wavefront produced by a point or line-source feed into a planar or linear wavefront; i.e., they are focused at infinity. In practice, however, complex feeds can be used since performance does not deteriorate rapidly with small off-axis feed displacement. Other lens designs have limited application at microwave frequencies.

Figure 18-1 shows two ways of achieving lens designs; in methods *a* and *b* rotational lenses are used in conjunction with a point-source feed (the term *rotational* means that the surfaces of the lens are obtained by revolution of a curved line about the lens axis), while in methods *c* and *d* cylindrical lenses (i.e., lenses whose surfaces are generated by moving a curved line perpendicularly to itself) are used in conjunction with a line-source feed.

At microwave frequencies, natural homogeneous dielectric media always have $n > 1$ (i.e., a phase velocity less than that of light in free space), which leads to the convex lens shapes of Figure 18-1*a* and *c*. However, artificial or fabricated media may be constructed with a range of n from $n \gg 1$ to $n \ll 1$. Examples of lenses constructed with $n < 1$ are shown in Figure 18-1*b* and *d*, where it is seen that a concave shape is required to focus a beam. In general, the types having $n < 1$ are highly dispersive (i.e., n varies rapidly with frequency), while those having $n > 1$ are nondispersive. Thus lenses having $n < 1$ are usually limited to small frequency bandwidths, while lenses having $n > 1$ may be designed to operate over an octave or more.

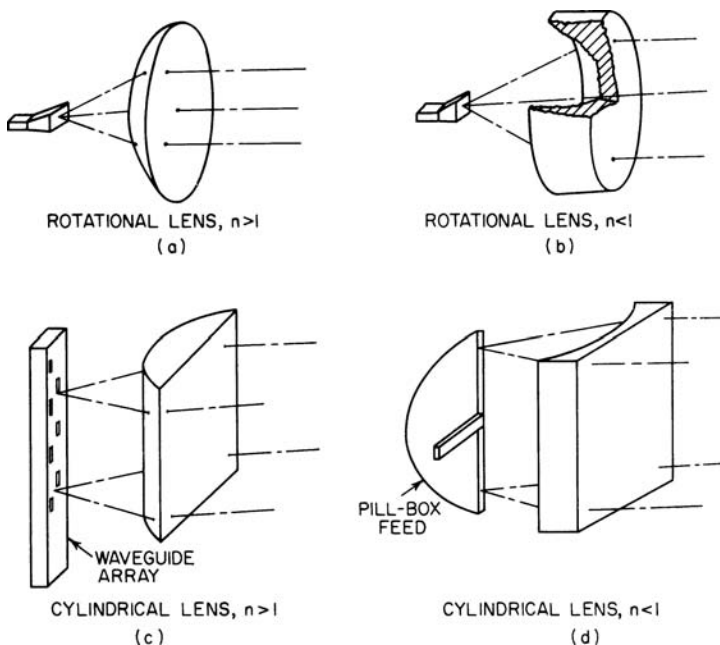


FIGURE 18-1 Basic lens configurations

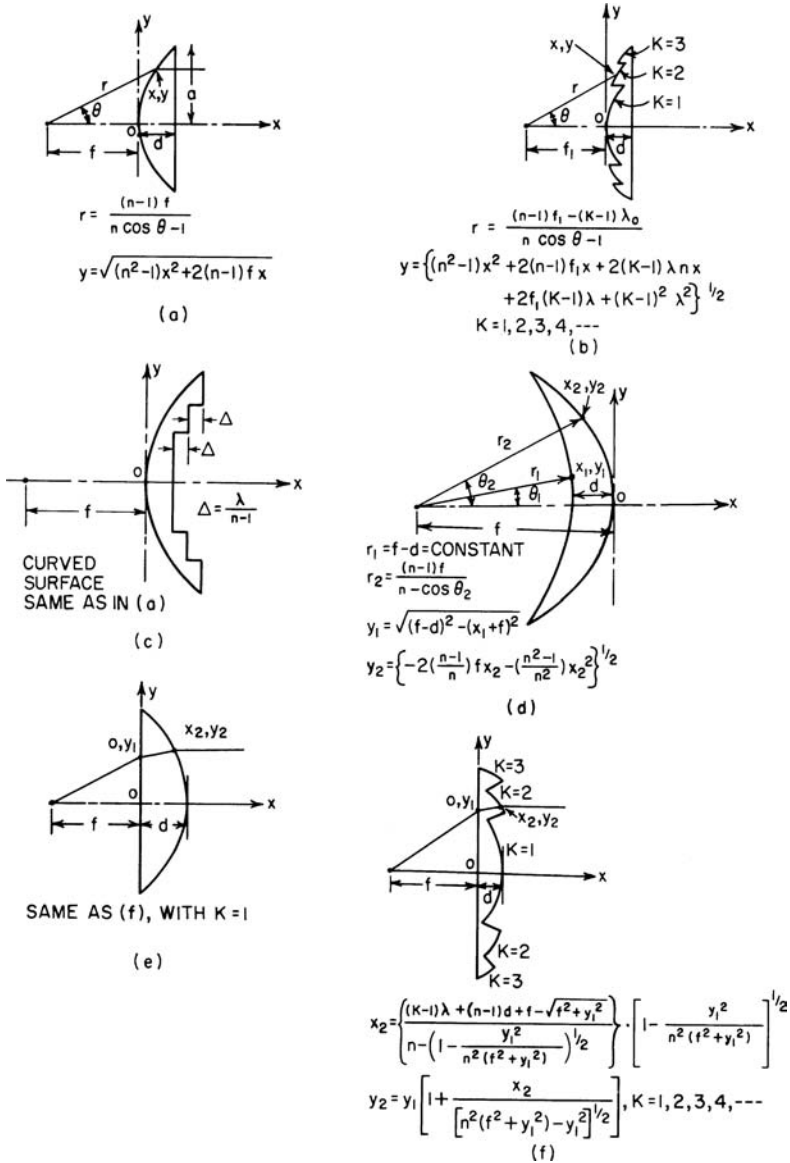
18.2 LENS-SURFACE FORMULAS FOR $n > 1$

Common shapes for $n > 1$ media lenses are shown in Figure 18-2; all are single-degree-of-freedom, single-focal-point lenses because either they refract at only one surface, or one surface is fixed during the design. The formulas given in the figure apply to both rotational and cylindrical surfaces.

Single Refracting Surface, $n > 1$

The lenses in Figure 18-2a–d refract at only one surface. In lenses a–c, refraction occurs at the inner surface (adjacent to the focal point) and in lens d, at the outer surface. In each case, the nonrefracting lens surface is parallel to a wavefront. The lens-surface formulas in Figure 18-2 are derived by equating the optical-path length of a general ray to that of the central ray from an on-axis point at a distance f from the inner surface to a planar wavefront perpendicular to the axis; for example, for the lens in Figure 18-2a, the relation is $r = f + n(r \cos \theta - f)$, which can be easily manipulated into the form given in Figure 18-2a, a hyperbola. The feed is located at the focal point farthest removed from the inner surface.

A disadvantage of this type of lens is that reflections from the nonrefracting surface will converge at the focal point, since the surface is coincident with a wavefront, and cause a feed-line mismatch approximately equal to the surface-reflection coefficient. It is desirable that this surface reflection be either prevented from entering the feed line or eliminated. A high feed-line voltage-standing-wave-ratio (VSWR) may be avoided by using a ferrite


 FIGURE 18-2 Lens-shape designs for $n > 1$ media

isolator, by tilting the lens slightly (as in Figure 18-3a) so that the reflected energy will be refocused off the feed, or by displacing half of the lens a quarter wavelength (in the lens medium) along the axis with respect to the other half (as in Figure 18-3b) so that reflections from the two lens halves are 180° out of phase and create a null at the feed center. These methods have been found to be effective in reducing the feed-line VSWR; however, they

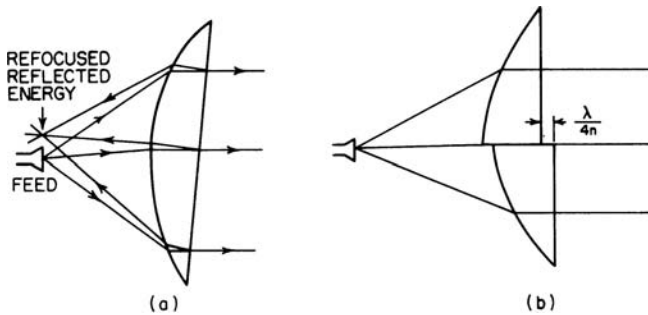


FIGURE 18-3 Reduction of feed VSWR by (a) lens tilt and (b) quarter-wavelength displacement of half of the lens

do not eliminate the other effects of surface reflection, such as loss of gain and increase in sidelobe level. Whenever these factors are important, the surface-matching techniques to be described later are recommended. Avoiding feed-line mismatch by the technique of Figure 18-3a depends on low wavefront distortion for small off-axis feed displacements. Analysis⁷ of aberrations due to feed displacement from the focal point shows that coma is the predominant term that limits performance, that coma is reduced as focal length and n are increased, and that the coma for a large- n hyperbolic lens is twice that for an equivalent-focal-length paraboloidal reflector. Although this is not usually necessary for small lenses, if the technique of Figure 18-3b is used for large-diameter lenses, the two lens halves should be designed with different focal lengths.

These lenses may be constructed with natural dielectric materials and artificial delay media. Kock⁸ developed a number of artificial delay media and constructed lenses of the type shown in Figure 18-2a and b with these media.

Two Refracting Surfaces, $n > 1$

Microwave lenses that refract at both surfaces have not been used as frequently as the single-refracting-surface types because their performance is not sufficiently better to justify the more difficult design problems.

Lens designs that meet the Abbe sine condition (to eliminate coma for small off-axis feed displacements) have been developed.⁹ The planoconvex single-focal-point lenses in Figure 18-2e and f almost satisfy the Abbe sine condition for $n = 1.6$ and an $f/D = 1$. An experimental 50-wavelength-aperture lens had good performance over a total scan angle of 20° .

A lens with two refracting surfaces has two degrees of freedom that should permit designs with two focal points. Several investigators¹⁰⁻¹² have developed bifocal design techniques that are iterative and do not lead to closed-form solutions. These lens designs are two-dimensional (cylindrical) and symmetrical about the lens axis. One technique¹⁰ assumes prisms at the lens edges and represents surfaces with even-order polynomials to permit iterative ray tracing. Another technique¹¹ involves a lattice method wherein points on the two lens surfaces are determined by alternating between the two focal points for ray tracing. Brown¹² uses prisms at the lens edges and the lattice method for determining the surfaces. For $n = 1.6$ and focal points at $\pm 20^\circ$, the resulting lens is almost planoconvex (the almost linear surface is slightly concave), very nearly that obtained in Sternberg.¹⁰ Zoning is achieved by designing separate lenses for each of the zones. Three-dimensional lenses were obtained by rotating the two-dimensional designs about the lens axis. In these

lenses, astigmatism is large and limits the beam-scan angles to about 20° . Brown¹² shows experimental data for 50-wavelength-diameter lenses with $n = 1.6$ and focal points at $\pm 20^\circ$ for f/D 's from 1 to 2.

Zoning Formulas, $n > 1$

In the case of physically large lenses, the use of continuous surfaces (as in Figure 18-2a, *d*, and *e*) results in a massive structure that is difficult to produce without imperfections, and the long path lengths in the medium make its index of refraction highly critical. To alleviate these disadvantages, use is commonly made of discontinuous, or zoned, surfaces as in Figure 18-2b, *c*, and *f*. Ray paths through any two adjacent zones are designed to differ from each other by exactly 360° (or a whole multiple thereof) at the design frequency so that a plane wavefront results on the outer side of the lens.

Figure 18-2b and *c* gives the surface-shape formulas for zoned versions of the lens of Figure 18-2a, while Figure 18-2f shows the zoned counterpart of Figure 18-2e; the use of integers $K = 1, 2, \dots$ gives a family of lenses that has a focal point at $x = -f$ and produces plane wavefronts that differ from each other in phase by integral multiples of 360° . In the use of these formulas, a thickness d on the axis is chosen and the central zone of the surface is computed for $K = 1$. At the radius for which the lens thickness is equal to the minimum allowed by mechanical considerations, the second zone is computed for $K = 2$. In a similar manner, additional zones are computed with $K = 3$, etc., until the desired lens diameter is attained. If the minimum allowable thickness is d_{\min} , the maximum thickness is approximately $d_{\min} + \lambda/(n - 1)$, where λ is the design wavelength.

Effects of Zoning on Aperture Illumination, $n > 1$

The aperture illumination of a lens is determined by a number of factors, one of which is the lens-material loss that is proportional to lens thickness. In an unzoned lens with $n > 1$, this loss is larger for on-axis rays where the lens has maximum thickness than for edge rays. This leads to a more nearly uniform amplitude distribution than would be estimated without considering loss. Since a zoned lens is thinner and therefore has less loss, the aperture illumination is less strongly affected.

Certain defects in the aperture illumination usually occur with zoned lenses and should be considered carefully in the lens design. For example, Figure 18-4a shows how shadowed bands (without energy) occur in the aperture field of a refracting outer surface. Inspection of the figure reveals that there is no way in which the zonal boundary can be shaped to eliminate the shadow. These nonilluminated bands produce aperture illumination discontinuities that increase sidelobe levels and decrease gain.

A second type of defect occurs when a first refracting surface is zoned, as in Figure 18-4b. In this case, the rays between rays *a* and *b* are not properly refracted by the lens and are scattered in undesired directions. Although there is no shadowing in this case, the aperture illumination is perturbed at the zone boundaries and deteriorates the expected radiation pattern.

However, zoning without shadowing or energy loss is possible if it is done on a nonrefracting surface of the lens. Figure 18-4c shows an example for the case of a plane outer surface. The surface of the step between zones must be perpendicular to the zoned surface. This type of zoning should provide the most satisfactory aperture illumination, but disturbance still exists along the zone boundaries because of phase differences between the rays just inside and outside the dielectric surface. In the special instance of a cylindrical lens in which the **E** vector is everywhere perpendicular to the step surface, this disturbance may be avoided by covering the zone boundaries with thin conducting sheets.

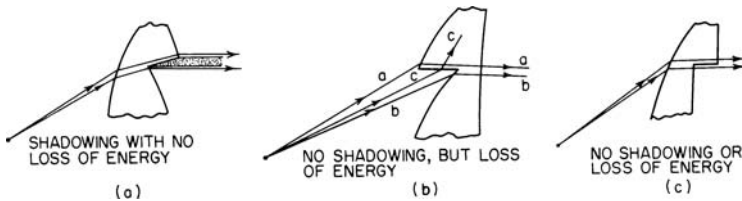


FIGURE 18-4 Effects of zoning on aperture illumination, $n > 1$

Bandwidth of Zoned Lenses, $n > 1$

The surface-shape design of an unzoned constant- n lens is independent of frequency. However, as may be seen from the formulas in Figure 18-2, zoning introduces a frequency dependence so that a zoned lens has a limited bandwidth of satisfactory operation. Rays through adjacent zones will differ by exactly 360° only at the design frequency. If the lens has a total of N wavelength steps between the central and edge zones, the effective path-length difference between a ray through the edge zone and a ray through the central zone is $N\lambda_0$ at the design wavelength λ_0 and $N\lambda$ at any other wavelength λ , where N is 1 less than the K of the outermost zone. The phase difference, in wavelengths, of the edge zone compared with the central zone is $-Nd\lambda$, where $d\lambda = \lambda - \lambda_0$. Since the total bandwidth is approximately $100(2d\lambda/\lambda)$ percent, if a phase difference of $\pm\lambda/8$ is allowed, the bandwidth of a lens having N wavelength steps is $25/N$ percent. Thus a zoned lens has an approximate bandwidth of 5 percent for $N = 5$ if only zoning effects are present. However, since there are other sources of phase error and it is necessary to allocate the total allowable phase error among all error sources, the bandwidth may be smaller than indicated earlier.

18.3 FACTORS AFFECTING GAIN AND SIDELobe LEVELS OF LENS ANTENNAS

The gain of a lens antenna is reduced by a number of factors that include the amount of feed energy not incident upon the lens (spillover loss), the effects of lens shape on aperture illumination, dissipation loss in the lens medium, and reflection from the surfaces.

Spillover Loss

Spillover may be avoided if the feed horn is extended to the lens edges; however, this is usually not done, since lens-surface reflections are reflected by the horn walls and radiated from the lens in undesired directions, thereby increasing sidelobe levels. The horn walls can be covered with absorbing material to reduce the energy in the reflected waves, but with this technique the antenna gain is reduced since feed energy is also absorbed.

In applications in which low sidelobe levels are important, the use of a multielement (array) feed to tailor the aperture illumination and reduce spillover is recommended. Also, surface matching is recommended to permit the use of an extended horn, increase gain, and reduce sidelobe levels.

Effect of Lens Shape on Aperture Illumination

The aperture amplitude distribution is a function of the feed radiation pattern, the shape of the lens, lens-surface reflections, and lens losses. Jasik¹ and Silver² derive the relations for conversion from the power per solid angle radiated by the feed to the power per unit area in the aperture for several lens types and plot the conversion factors for several sets of lens parameters. In general, for $n > 1$ lenses, the conversion factor is < 1 ; for $n < 1$ lenses, the factor is > 1 ; therefore, if a given feed is used for the two lens types, the $n < 1$ lens will have a higher edge illumination than the $n > 1$ lens.

Dissipation Loss in Lens Media

The dissipative attenuation constant for a dielectric medium is approximately $27.3 (\tan \delta) n$ (dB/wavelength), where n is the index of refraction and $\tan \delta$ is the loss tangent of the medium. Since the maximum thickness of a zoned $n > 1$ lens is approximately $\lambda/(n - 1)$, the upper limit for lens attenuation is approximately $27.3 (\tan \delta)n/(n - 1)$ (dB). Therefore, for most practical materials the maximum dissipation loss is several tenths of a decibel. Von Hippel¹³ contains extensive dielectric-constant and loss-tangent data for many dielectric materials over a large range of frequencies and temperatures.

Reflections from Lens Surfaces

The air-to-dielectric interface at each lens surface produces reflected and transmitted waves. The amplitude and phase of waves are obtained by the application of Fresnel's equations and depend on n , the angle of incidence, and the polarization relative to the incidence plane (a plane through the incident ray and the surface normal). Since there are reflections at both lens surfaces, the effects of internal multiple reflections are determined by the ray path lengths between the surfaces. The thickness of a typical lens varies appreciably over its aperture so that, with the transmitted waves adding in various phases, there are a number of almost sinusoidal variations in amplitude and phase over the aperture. For most lenses, incidence angles are less than 45° so that reflection losses can be averaged over all incident polarizations and angles to obtain an approximate lens reflection loss of $8.69(n - 1)^2/(n + 1)^2$ dB. For an $n = 1.6$ lens, this reflection loss is less than 0.5 dB. However, sidelobe levels are seriously increased and for most lenses are limited to values of about -20 dB.

These effects may be reduced by matching the lens surfaces.¹⁴⁻¹⁶ For near-normal incidence angles, this can be accomplished by adding to each surface a quarter-wavelength-thick coating (in the material) with an index of refraction of $(n)^{1/2}$, where n is the lens-material index of refraction. If it is necessary to match for nonnormal incidence angles, the coating thickness and index of refraction may be modified as discussed in Jasik.¹

For many lens materials, a coating material with the correct index may not be available; in that case, surface matching may be performed by machining the lens surfaces to the configurations shown in Figure 18-5 and discussed in detail in the literature.^{1,14-16} The improvement that can be obtained in lens performance by lens-surface

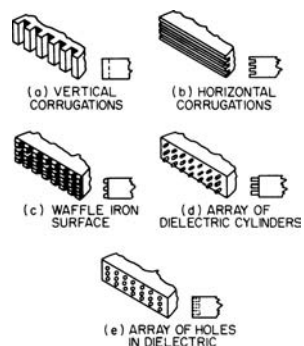


FIGURE 18-5 Simulated quarter-wavelength matching transformers for lens surface

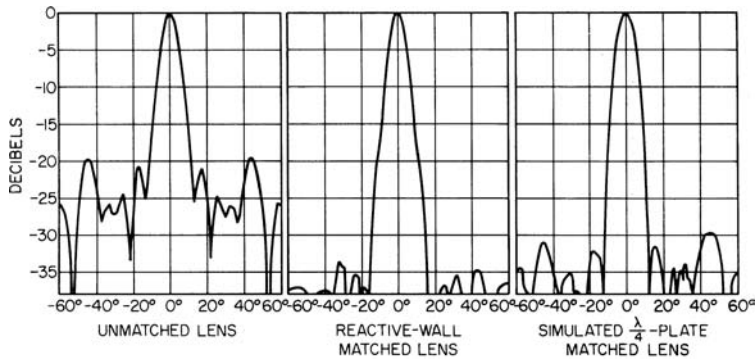


FIGURE 18-6 Radiation patterns of unmatched and matched lenses

matching¹⁶ is shown in Figure 18-6; however, these lenses were matched with artificial dielectrics not discussed in this chapter.

18.4 LENS-SURFACE FORMULAS FOR $n > 1$

The lens formulas given in Figure 18-2 apply for $n < 1$ as well as for $n > 1$; however, for convenience, the lens-surface formulas for $n < 1$ media planoconcave lenses are shown in Figure 18-7 and apply to both rotational and cylindrical surfaces. These are single-refracting-surface, single-focal-point lenses since the planar outer surface is fixed during the design to lie on a wavefront. The inner surface is an ellipse with foci on the axis; the feed is located at the focus farthest removed from the inner surface. Reflections from the outer surface converge at the focal point as for $n > 1$ lenses and can be eliminated by techniques used for those lenses.

For the stepped inner surface in Figure 18-7b, shadow bands at the steps cannot be avoided, as may be seen in Figure 18-8. Although all the rays emanating from the feed are collimated by the lens, the steps cause discontinuities in the aperture illumination. As in the case of $n > 1$ lenses, zoning can be accomplished without shadowing or energy loss only if the steps are formed in the equiphase outer surface, but the resulting meniscus-type lens shape is more difficult to manufacture and to support than the flat shape of Figure 18-7b.

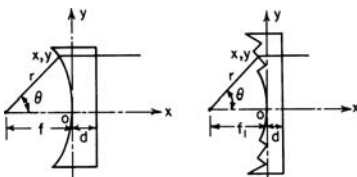


FIGURE 18-7a and b Lens-shape for $n < 1$ media

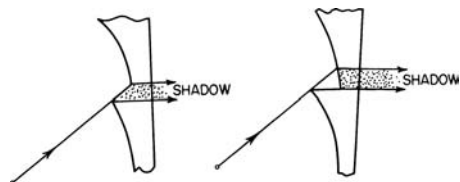


FIGURE 18-8 Shadows introduced by zoned lens in Figure 18-7b

The maximum thickness of the stepped lens for $n < 1$ is equal to $d_{\min} + \lambda/(1 - n)$, where d_{\min} is the minimum thickness permitted in the mechanical design. The bandwidth limitation due to zoning alone is the same as for $n > 1$ lenses. However, the bandwidth of the actual lens is considerably less than this when the frequency sensitivity of n is taken into account, since all known media having $n < 1$ exhibit a large rate of change of n with frequency. Thus, in the common case of a metal-plate zoned lens having the shape shown in Figure 18-7*b*, if $\pm\lambda/8$ phase error is allowed, the bandwidth is approximately $25n_0/(1 + (N + 1)/n_0)$ percent, where n_0 is the n at the design frequency and N is the number of one-wavelength steps between the central and edge zones. In the case of frequency-sensitive media, zoning increases the bandwidth of a lens over that of an equivalent unzoned lens by a factor of 2 or 3 because the ray path lengths in the dispersive medium are greatly reduced by zoning. The bandwidth of a given lens antenna may be extended considerably if the antenna application permits moving the feed position along the lens axis as a function of frequency, since defocus is the principal effect of a frequency change.

The lenses in Figure 18-7 may be constructed with artificial dielectric, metal-plate, or waveguide media. An excellent summary of artificial dielectric media is contained in Jasik.¹ Metal-plate and waveguide media are discussed in the next section. Kock¹⁷ used a number of techniques to construct metal-plate lenses with both circular and rectangular apertures. Figure 18-9 is a photograph from Kock⁸ of a stepped 96-wavelength-aperture, $F/D = 0.96$, metal-plate lens that has a useful bandwidth of about 5 percent. Most early metal-plate lenses used the designs and techniques of Kock. Additional $n < 1$ lenses are discussed in Section 18.6.

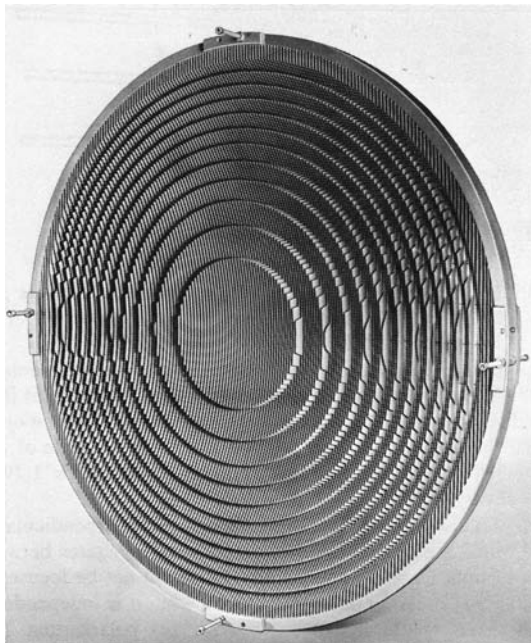


FIGURE 18-9 An $F/D = 0.95$ zoned lens

18.5 METAL-PLATE AND WAVEGUIDE MEDIA

Rays incident on natural dielectric and many artificial dielectric media obey Snell's law at the surface. A metal-plate medium (composed of parallel, equally spaced plane metal sheets) constrains rays in the medium to paths parallel to the plates; therefore, Snell's law is not obeyed for incident rays in a plane normal to the plates (see Figure 18-10), while it is obeyed for rays incident in a plane parallel to the plates (see Figure 18-11a). A waveguide medium constrains rays in the medium to paths parallel to the waveguide axis (see Figure 18-11b); therefore, Snell's law is not obeyed for any incidence angle.

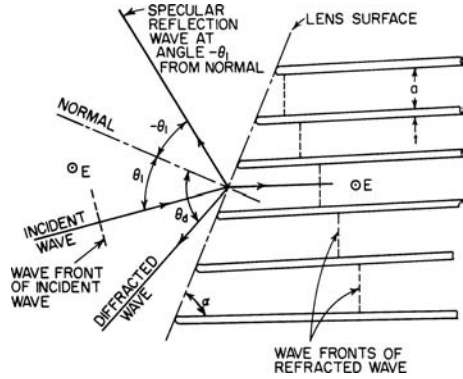


FIGURE 18-10 Constrained refraction in the H plane

Both metal-plate and waveguide media operate in the fundamental TE mode of propagation, and since the phase velocity exceeds the velocity of light in free space, $n < 1$. If the cutoff wavelength is λ_c , then $n = [1 - (\lambda/\lambda_c)^2]^{1/2}$ for an operating wavelength λ . For both parallel plates spaced a apart and a waveguide of width a , $\lambda_c = 2a$. For a regular hexagon of width a between parallel sides, $\lambda_c = 1.792a$. For a circular waveguide of diameter a , $\lambda_c = 1.705a$.

When parallel plates are used, an E -field component perpendicular to the plates must be avoided, since this excites a TEM mode that propagates between the plates with the velocity of light ($n = 1$) and the energy would not be focused. For square, circular, and

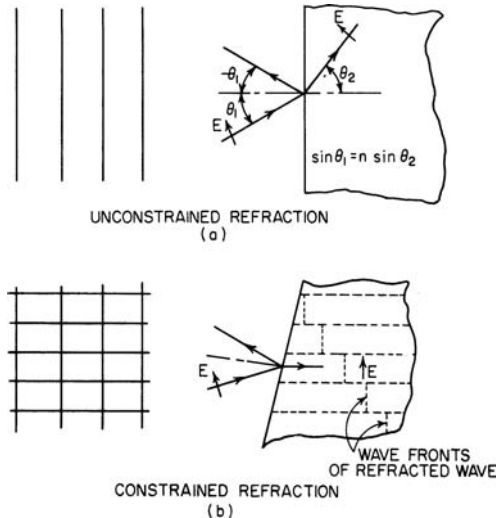


FIGURE 18-11 Unconstrained (a) and constrained (b) refraction in the E plane

hexagonal cross-section waveguide media, n is independent of incident polarization, and these structures can be used for any polarization. A rectangular cross-section waveguide medium has different values of n for the two principal linear polarizations.

Limitations on Metal-Plate Spacing and Rectangular-Waveguide Width

A basic limitation on the choice of plate spacing a results from the fact that the medium must operate between the cutoff frequencies of the fundamental and the next higher TE modes. For air-filled structures, this requires that $0.5 < a/\lambda < 1.0$; this range of a/λ corresponds to a range in n of $0 < n < 0.866$. However, a further restriction on the plate spacing must be imposed if the existence of diffracted waves due to the grating effect is to be avoided. It will be shown later how such diffracted waves cause considerable loss of transmitted power, distortion of the aperture illumination function, and consequent loss of gain and increased sidelobe levels. The presence of one diffracted wave is shown in Figure 18-10. This occurs in the direction in which the diffracted waves from the individual plate edges combine in phase. It can only occur when the plate spacing exceeds a particular value and will be avoided if the following condition is met:

$$[\lambda \sin \alpha / (a + t)] - 1 > |\sin \theta_1| \quad (18-1)$$

where t is the wall thickness and the other symbols are defined in Figure 18-10. This is also the condition for avoiding a diffracted wave when the incident wave is in the medium and the transmitted ray emerges at angle θ_1 . For $\alpha = 90^\circ$, this is the familiar expression for avoiding grating lobes from a phased array: $s = a + t < \lambda / (1 + \sin \theta_1)$. The limiting values of Eq. 18-1 are shown graphically in Figure 18-12, where the angle θ_1 of the ray in free space is plotted versus $(a + t)/\lambda$ for different values of α from 30 to 90° . For each particular value of α , the region to the lower left of the curve gives the ranges of θ_1 and $(a + t)/\lambda$ in which diffracted waves are avoided.

Reflection and Transmission at an Interface

For the case of the E field parallel to the plane of incidence and parallel to the plates (see Figure 18-11a), the power-reflection coefficient is

$$r^2 = \sin^2 (\theta_1 - \theta_2) / \sin^2 (\theta_1 + \theta_2) \quad (18-2)$$

where θ_1 is the ray angle in free space and θ_2 the ray angle in the parallel-plate medium, both angles being measured from the normal to the surface. The plates are assumed to be very thin; however, thick plates are treated in Primich¹⁸ for $\theta_1 = 0$ and $\alpha = 90^\circ$. Equation 18-2 also holds true for a wave incident from the metal-plate region. The reflected ray is then within the metal-plate region at an angle $-\theta_2$. Note that Eq. 18-2 also applies for a solid dielectric with E perpendicular to the plane of incidence. Equation 18-2 and Snell's law of refraction hold for all angles of incidence [if $0.5 < (a + t)/\lambda < 1.0$], and diffracted waves do not occur. However, this is not true of parallel polarization incident on a constrained medium (see Figure 18-11b), and a solution for that case is unavailable.

Considerable published information is available for the case of E perpendicular to the plane of incidence and parallel to the edges of a set of very thin plates (see Figure 18-10). The basic analysis is due to Carlson and Heins,¹⁹ while further theoretical work was done

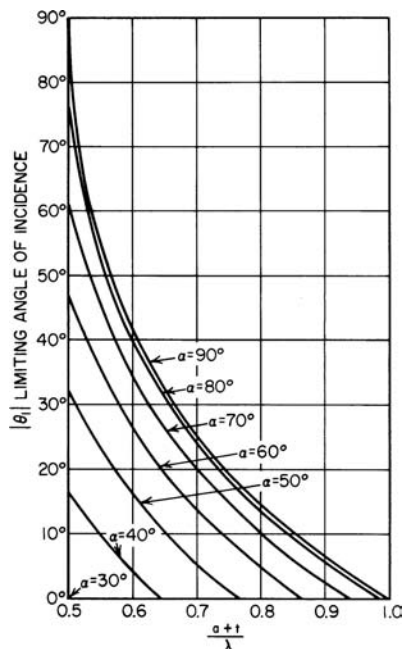


FIGURE 18-12 Limiting angle of incidence for pure specular reflection versus plate spacing

by Lengyel,²⁰ Berz,²¹ and Whitehead.²² In the range of θ_1 free of diffracted waves, as given by Eq. 18-1 and Figure 18-12, the power-reflection coefficient for a wave incident on an array of plates is

$$|r|^2 = \left[\frac{\cos(\theta_1 + \psi) - n}{\cos(\theta_1 + \psi) + n} \right] \left[\frac{\cos(\theta_1 - \psi) - n}{\cos(\theta_1 - \psi) + n} \right] \quad (18-3)$$

where $\psi = 90^\circ - \alpha$ is the angle between the normal to the boundary and the plates. The specularly reflected wave propagates at angle $-\theta_2$. Because of reciprocity, this formula also applies to waves in the plate region incident upon the free-space boundary, where the waves between the plates are phased to produce a transmitted wave propagating at angle θ_1 from the normal. For normal incidence and $\alpha = 90^\circ$, Eq. 18-3 reduces to $r^2 = (1 - n)^2 / (1 + n)^2$, which also applies to a solid-dielectric boundary. In the range of θ_1 where diffracted waves may exist, Eq. 18-3 is no longer valid. Formulas for this range have been derived by Lengyel,²⁰ Berz,²¹ and Whitehead²² and are much more complicated than Eq. 18-3.

Figure 18-13 shows the percentage of incident power transmitted into a metal-plate medium for the conditions of Figure 18-10 and $\alpha = 90^\circ$. To the left of the break in each curve (points D), the nontransmitted power is reflected in the specular reflected wave, while to the right of the break the nontransmitted power is divided between the specular reflected wave and a diffracted wave. The power loss in the latter region is large and should be avoided in lens design. At points S, the angle θ_d of the diffracted wave is equal to θ_1 of the incident wave (see Figure 18-10). When $\alpha = 90^\circ$, the reflection and transmission quantities are not symmetrical about $\theta_1 = 0$.

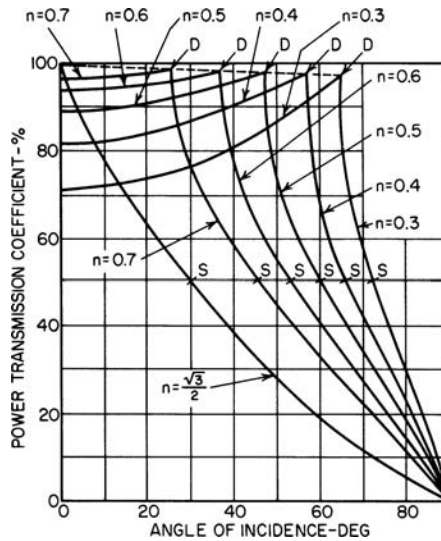


FIGURE 18-13 Percentage of power transmitted into metal-plate medium, $\psi = 0$

The discontinuity effect at the boundary of the metal-plate medium also results in a phase change. Formulas giving this change have been derived by Lengyel,²⁰ Berz,²¹ and Whitehead.²² Graphical plots of this phase change versus the angle of incidence are given by Berz and Whitehead and show a moderate variation with the angle of incidence. Calculations by Whitehead for a typical planoellipsoidal lens show that substantial curvature of the wavefront occurs at the aperture when the feed element is at the focal point but that the wavefront may be made almost plane by moving the element slightly away from the focal point.

The just-described data for perpendicular polarization incident on a set of parallel plates apply equally well to an array of thin-walled square or rectangular tubes (see Figure 18-11*b*). However, for parallel polarization the constraint offered by the array of tubes has a large effect on its reflection characteristics. Wells²³ shows that, at the angle of incidence for which the diffracted wave first appears, a sharp peak of specular reflection also occurs.

Methods of Reducing Reflection Losses

Within the angle-of-incidence range given by Eq. 18-1 and Figure 18-12, the reflection from the surface of a metal-plate or waveguide lens may be canceled through the introduction of an additional reflecting discontinuity. Possible structures are shown in Figure 18-14, where in *a* a dielectric sheet is placed outside the lens and in *b* obstacles are placed inside the lens. In both cases the magnitude and phase of the additional reflection are adjusted to cancel the reflection from the plates. The magnitude depends upon the dielectric constant and the thickness of the dielectric sheet or the shape and size of the obstacles, while the proper phase relationship is obtained through choice of the spacing of the dielectric sheet or obstacles from the plate edges. The correct design for a given lens may be obtained experimentally; however, a theoretical design may also be computed through the use of graphical and tabulated data given by Lengyel²⁰ for the magnitude and phase of the reflection coefficient of the boundary.

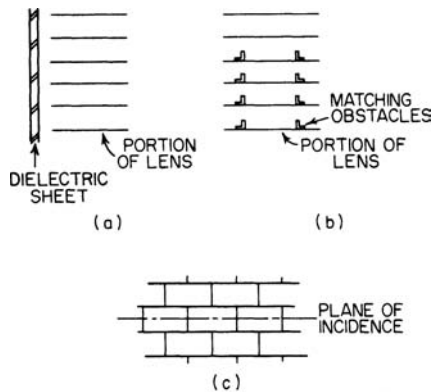


FIGURE 18-14 Techniques for reducing surface reflections from metal-plate lenses

In a practical lens design, it is difficult to avoid the ranges of θ_l for which diffracted waves occur. Many early designs of metal-plate lenses did not do this, and thus considerable loss of gain resulted. One method of avoiding this condition, as may be seen from Figure 18-12, is to reduce the plate spacing; however, this reduces n and increases the reflection loss at angles near normal incidence and reduces the bandwidth. A further possibility is to fill the regions between the plates with dielectric material or to introduce ridges in the waveguide parallel to the axis (see Figure 18-14b) in a manner analogous to ridge waveguide.^{24,25} In this way, the plate spacing may be reduced to a point at which diffraction is no longer a problem while n remains at a reasonable value.

Another method of avoiding diffracted waves in one plane is to use a staggered arrangement of waveguide channels,²³ as in Figure 18-14c, now a proven technique in space-fed phased arrays. The spacing between scattering edges is effectively cut in half, thus doubling the frequency at which a diffracted wave first appears. This staggered arrangement also eliminates the sharp peak of specular reflection that occurs with the nonstaggered array of tubes for parallel polarization.

18.6 WAVEGUIDE LENSES

Three-dimensional unzoned waveguide lenses of the type shown in Figure 18-7a have been designed. The inner surface is an ellipsoid of revolution about the lens axis, and the outer surface is a plane perpendicular to the axis. One such lens²⁶ has a 36-in (914-mm) diameter and a 31.2-in (792-mm) nominal focal length and is constructed of 0.68-in- (17.3-mm-) inside-diameter square waveguide to provide a nominal $n = 0.5$ at 10 GHz. The lens was fed with a circular-aperture-waveguide horn that produced circular polarization. The reference has a number of plots of experimental data: gain, beamwidth, n , and focal length versus frequency; gain, beamwidth, and first sidelobe level versus feed-horn diameter; radiation patterns as a function of beam-scan angle; and gain, beamwidth, and first sidelobe level versus beam-scan angle. The sidelobe level increases from -22.5 dB at 13° beam scan and remains at about that level for beam scan to 22° . The gain decreases by 3 dB (from the on-axis value of 36 dB) at a beam-scan angle of 18° , or 7.65 beamwidths.

Another design²⁷ has a 24-in (610-mm) diameter and a 36-in (914-mm) focal length constructed of $5/_{32}$ -in (4.0-mm) cell honeycomb to provide an $n = 0.642$ at 55 GHz. Because the honeycomb cells deviated from perfect hexagons, the performance varied with the linear-polarization orientation. The aperture efficiency was only 18.3 percent because random variations in the cell size produced large aperture phase variations.

To obtain wide-scan-angle performance, Ruze²⁸ investigated bifocal cylindrical constrained lenses in which the two focal points are equally displaced from the lens axis at P_1 and P_2 , as shown in Figure 18-15. The inner surface is found by equating optical-path lengths for the central ray and a general ray from each of the focal points. It is an ellipse, with foci at the two focal points, independent of n or lens thickness, given by

$$(x + af)^2 / (af)^2 + y^2 / f^2 = 1 \quad (18-4)$$

where $a = \cos \phi$. Imposing the two focal points uses two of the three degrees of freedom available. Ruze considered several alternatives for choosing the third condition: constant n , constant thickness, linear outer surface, third correction point, and no second-order error; for the last four, n varies with y . The deviations from a linear phase front for a general feed point were expanded into a power series and examined for each case. For the on-axis feed point there are no odd-order terms, and the second-order deviation (defocus) can be corrected by moving the feed point along the axis; however, this correction is not needed for all cases. The feed arc is chosen as a circle through the corrected on-axis feed position and the two focal points. Several of these designs permit a beam scan of 110° for $\phi = 50^\circ$. Feed-arc selection for reduced phase errors is discussed in Peebles²⁹ and in Rappaport and Zaghoul.³⁰

The constant- n lens has been the most popular. The thickness is given by

$$d = d_0 + (m\lambda + ax) / (n - a) \quad (18-5)$$

where m is the number of wavelength steps in the outer surface. Proctor and Rees³¹ analyzed this type of lens with results near those of Ruze. The constant- n lens was expanded to three dimensions by Fine and Reynolds;³² the inner surface is an ellipsoid of revolution about the

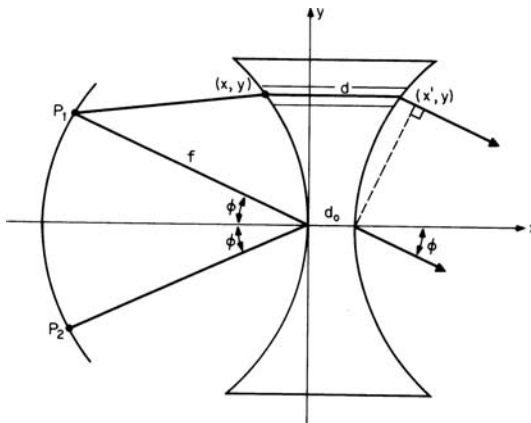


FIGURE 18-15 Bifocal metal waveguide lens

line through the focal points, and the outer surface is an ellipsoid with unequal principal radii. The performance is well behaved for feed points between the two focal points, but performance deteriorates rapidly for feed points beyond the focal points. Performance also degrades rapidly as the feed point is moved away from the xy plane.

A constant- n lens was used in an application³³ for which the scanning feed system was most conveniently designed for a planar feed aperture. The phase errors were minimized for the on-axis feed point by a method of stepping the outer surface. A constant- n lens was also used for a satellite antenna,³⁴ but with the two correction points collapsed to a single on-axis point so that the inner surface was a segment of a sphere and the outer surface a segment of a spheroid. The effects of zoning the outer surface were computed. The radiation-pattern effects of zoning were also computed in Lu.³⁵

A linear-outer-surface two-dimensional lens between parallel plates was constructed³⁶ for an X-band multibeam application by milling waveguides of unequal width (to obtain variable n with y) in metal blocks. The lens had good performance, and the tolerances obtained appeared sufficient to permit this technique to be used for frequencies up to 40 GHz.

Waveguide lenses have been designed to provide a larger bandwidth by designing for equal time delay^{37,38} for all rays through the lens. The phase delay in each waveguide is adjusted with a phase shifter to provide a planar wavefront. These lenses have a spherical inner surface so that they satisfy the Abbe sine condition for off-axis performance. Dion³⁸ has both single- and dual-frequency designs that obtain bandwidths that vary from 40 to 20 percent for lens diameters from 20 to 100 wavelengths.

18.7 BOOTLACE-TYPE LENSES

Bootlace lenses can have four degrees of freedom that permit designs with four focal points to obtain wide-angle performance for either multiple simultaneous beams or a scanning beam; however, most designs fix the outer surface and have three focal points. These constrained lenses have radiators on the inner and outer surfaces connected by TEM-mode transmission lines (usually coaxial or stripline), which give the appearance of untightened bootlaces, hence the name. Typically, all rays in these lenses have equal time delays from a focal point to the corresponding linear wavefront, and therefore they are inherently broadband, limited only by component bandwidths and the performance of a fixed aperture size.

Gent³⁹ identified most of the useful properties of this type of lens, including (1) four degrees of freedom, (2) the fact that one surface may be displaced and/or rotated relative to the other (i.e., the surfaces may have independent coordinate systems), (3) the ability to add phase shifters, attenuators, and/or amplifiers in the lens transmission lines to provide outer-surface aperture illumination control and beam scan, (4) the fact that the line lengths can be stepped by multiple wavelengths, analogously to the physical stepping shown in Figure 18-2 without creating shadowing but with reduced bandwidth, (5) the ability of radiators on the two lens surfaces to have different polarizations, and (6) the fact that the spacing between aperture radiators must be less than $\lambda/(1 + \sin \theta_{\max})$ to avoid grating lobes, where θ_{\max} is the maximum beam-scan angle from the aperture normal.

The Rotman-Turner⁴⁰ lens is a two-dimensional bootlace lens with a linear outer surface perpendicular to the lens axis and three focal points, one on axis and two equally displaced from the lens axis, shown in Figure 18-16 as points $G(-g, 0)$, $F_1(-\cos \phi, \sin \phi)$, and $F_2(-\cos \phi, -\sin \phi)$, respectively, where central rays from F_1 and F_2 subtend an angle ϕ with the axis. Desired wavefronts are displaced from the vertical by angles 0 , ψ , and $-\psi$, respectively, where $\psi = \phi$ in this design. All dimensions are normalized to the F_1 -to-origin distance. Pairs of inner- and outer-surface radiators are connected by coaxial lines with the general pair located at $P(x, y)$ and $Q(t, u)$, respectively. Let $w = s - s_0$ be the normalized

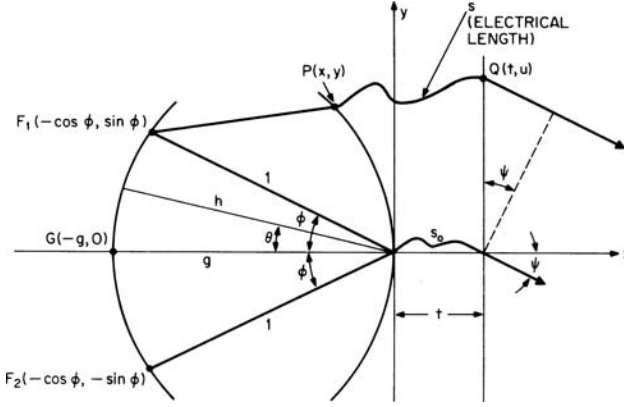


FIGURE 18-16 Rotman-Turner lens nomenclature

electrical path-length difference between the general and the central line lengths. Equating the optical-path lengths for a general ray and the central ray for each focal point produces

$$y = u(1 - w) \quad (18-6)$$

$$x^2 + y^2 + 2dx = w^2 + e^2u^2 - 2w \quad (18-7)$$

$$x^2 + y^2 + 2gx = w^2 - 2gw \quad (18-8)$$

where

$$d = \cos \phi$$

$$e = \sin \phi$$

x and y can be eliminated to obtain

$$aw^2 + bw + c = 0 \quad (18-9)$$

where

$$a = 1 - u^2 - (g - 1)^2/(g - d)^2$$

$$b = 2g(g - 1)/(g - d) - e^2u^2(g - 1)/(g - d)^2 + 2u^2 - 2g$$

$$c = ge^2u^2/(g - d) - e^4u^4/(4(g - d)^2) - u^2$$

w as a function of u is determined from Eq. 18-9 for fixed parameters ϕ and g . Substitution into Eqs. 18-6, 18-7, and 18-8 determines x and y . The technique of Ruze²⁸ (for the linear-outer-surface lens) for minimizing aberrations from a linear wavefront for a general feed point θ between $\pm\phi$ is used to determine $g = 1 + \phi^2/2$, for ϕ in radians. The feed path is chosen as a circle that intersects the three focal points.

Rotman and Turner⁴⁰ contains computations of w , x , and y as a function of u for $\phi = 30^\circ$ and computations of deviations from a linear wavefront for $0 \leq \theta \leq 40^\circ$ and several values of g .

Also shown are lens shapes for four values of g . For $\phi = 30^\circ$ and $g = 1.137$ (the optimum value of g for $\phi = 30^\circ$), the normalized path-length error is <0.00013 for $-0.55 \leq u \leq 0.55$ (i.e., aperture $= 2u_{\max} = 1.10$ and $g/2u_{\max} = 1.035$). If a phase error of $\pm\lambda/8$ is allowed, an aperture of 1000 wavelengths can be used to obtain a beamwidth $<0.075^\circ$ that can be scanned more than 60° , or more than 800 beamwidths.

An experimental lens at 3.0 GHz with $\phi = 30^\circ$, $g = 1.137$, $u_{\max} = 0.6$, and $D/\lambda = 18$ was constructed⁴⁰ by using RG-9/U coaxial cables for the lens, a TEM-mode parallel-plate structure between the feed position or positions and the inner lens surface, a flared parallel-plate aperture that contained the linear outer surface, and quarter-wavelength probes into the parallel-plate regions.

The Archer lens⁴¹ is a Rotman-Turner lens with additional features: (1) The parallel-plate region between the feeds and the inner surface is filled with a dielectric material of relative dielectric constant ϵ_r to reduce all linear dimensions by a factor of $n = (\epsilon_r)^{1/2}$, thereby permitting its use in a smaller volume. (2) Stripline or microstrip lens and feed-port connecting lines are printed on an extension of the dielectric material, thereby removing connectors and probes and the associated mismatches and permitting constructions to tight tolerances and wide-bandwidth operation. (3) The beam angle ψ can differ from the feed angle θ so that $\sin \psi / \sin \theta = K$ (where $K = 1$ for the Rotman-Turner lens)⁴²; $K > 1$ permits obtaining large beam-scan angles with practical feed positions, and $K < 1$ permits reduction of the lens size for limited scan-angle applications. These features increase the lens versatility, but do not increase the design degrees of freedom.

There is a nomenclature difference between the Archer and Rotman-Turner lenses. In the Archer lens, the printed-circuit lines that form the lens appear to be just connecting lines, while the dielectric region between the feed ports and the inner surface appears to be a lens and is usually referred to as such.

Archer lenses⁴¹ have been designed and constructed with materials with ϵ_r 's from 2.5 (Teflon fiberglass) to 233 (cadmium titanate). Figure 18-17 shows a 20-radiator, 16-beam lens constructed in microstrip on a barium tetratitanate ceramic substrate with $\epsilon_r = 38$. Most designs have a beam coverage of 120° , range from a 4-radiator, 4-beam lens to a 140-radiator, 153-beam lens, and have a typical operating frequency band of almost two octaves. Designs usually have adjacent beams overlap at the 3-dB points (with feeds at

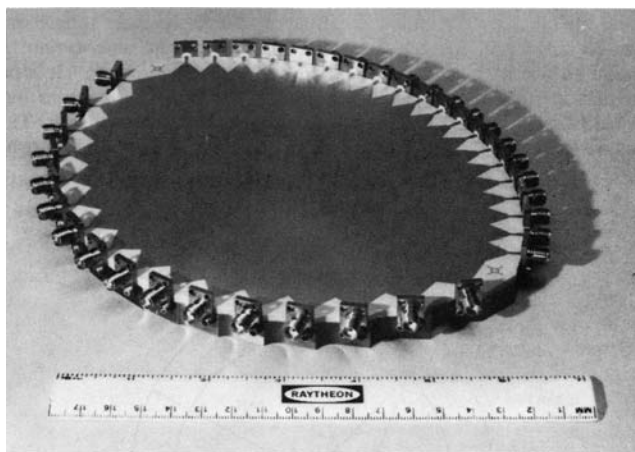


FIGURE 18-17 Archer lens, $\epsilon_r = 38$

equal intervals in sine space) at the maximum operating frequency where the gain and beamwidth are consistent with that obtainable from the full aperture. At lower frequencies, the beams broaden and overlap at a higher level, and gain decreases more rapidly than the beamwidth broadening would indicate owing to beam-coupling loss. Low sidelobe patterns⁴³ are obtained by feeding several adjacent beam ports with weighted amplitudes.

Alternative designs⁴¹ are obtained by fixing the outer surface as a circle, with an arc up to 60° , that is more adaptable to conformal applications. A circular aperture eases the design problem of eliminating aperture resonances that are common in linear arrays, increases the beam coverage with reduced gain roll-off with off-axis beam scan traded against on-axis gain, and reduces the effects of outer-surface mismatch since reflections do not have a constant phase difference from radiator to radiator as with a linear array.

The linear-aperture bootlace lenses discussed earlier produce fan beams. Multiple pencil beams in a plane can be produced by illuminating a parabolic-cylinder reflector⁴⁰ or cylindrical lens (see Figure 18-1) with the linear aperture. A two-dimensional cluster of pencil beams may be formed⁴¹ by stacking vertical linear-aperture lenses, so that the outer surface of each lens provides a column of radiators, and by feeding the beam ports of these lenses with the output ports of a stack of horizontal lenses. The beam ports can be used separately, adjacent beam ports may be combined to provide monopulse-type outputs, and multiple beam ports may be combined after weighting to provide low sidelobe beams. Three-dimensional bootlace-type lenses are discussed in Gent,³⁹ Archer,⁴¹ and in other literature.⁴⁴⁻⁴⁷

Gent³⁹ found a special case of the two-dimensional bootlace lens that has feed points and the inner surface on a circle of radius R and the outer surface on a circle of radius $2R$ with equal line lengths between the inner- and outer-surface radiators and recognized that it is equivalent to the R - $2R$ geodesic scanner.^{48,49} It is proved in Kales and Brown⁵⁰ that the R - $2R$ bootlace lens is unique. Bootlace R - $2R$ lenses have been constructed⁴¹ by using printed-circuit techniques on dielectric materials. The R - $2R$ type lens has been extended into an R - KR lens,⁴¹ shown in Figure 18-18, that has equal radial line lengths between radiators on a circle of radius KR and radiators on a circle of radius R . With $K = 1.92$, the optimum value, it is the printed-circuit version of a constant-index lens. It provides a nearly linear wavefront for an arc of

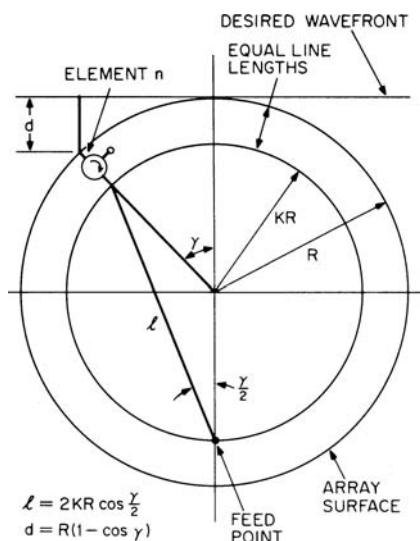


FIGURE 18-18 R - KR -array geometric relationship

approximately 120° on the circle with radius R . By adding a circulator in each of the lines to permit feed inputs, beams covering 360° can be obtained simultaneously.

18.8 PHASED-ARRAY LENSES

Phased arrays are discussed in Chapter 20; in this section space-fed phased-array lenses are discussed for comparison with other lenses.

Space-fed arrays⁵¹⁻⁵⁷ have constrained lenses that usually are constant-thickness^{51,54} with planar surfaces. Elements (or modules), usually perpendicular to the surfaces, contain inner- and outer-surface radiators, phase shifters, and in some cases other components such as phase-shifter drivers and polarization switches. The inner-surface radiator has the polarization of the feed or feeds and needs to be matched only for the incidence angles from the feed or feeds. The outer-surface radiator can have any polarization or polarizations that may be switchable, must be matched for the range of beam-scan angles used, and must withstand the environment. Most phase shifters are digital, modulo 360° , in which the number of bits is determined by the required beam-pointing accuracy and sidelobe levels. Either diode- or ferrite-type phase shifters are used, the type for a given application being determined by the operating frequency, radio-frequency power levels, required phase-state-switching speed, and switching and/or phase-state-holding power.

Beam steering and collimation (the conversion of the feed wavefront to a plane wave) are obtained by the phase shifters; the computed phase shifts required for the two functions are superimposed. The phase shift required for collimation is determined from ray tracing as for other lenses. Phase-shift commands may be computed and commanded on an element-by-element basis or by rows and columns.⁵² This collimation method reduces the number of phase-shifter bits required to obtain a given beam-pointing error, since the otherwise-present regularity of phase errors over the lens outer surface is destroyed.

In many pulsed-radar applications, it is possible to change the state of the phase shifters between transmit and receive; in these cases, separate transmit and receive feeds may be used. If sufficient isolation between feeds is obtained, the need for a transmit-receive switch is eliminated. In addition, separate feeds permit the lens illumination to be optimized for both transmit and receive; e.g., to obtain maximum gain on transmit and low sidelobes on receive.

For applications that require large beam-scan angles and large instantaneous-frequency bandwidth, lenses permit the total angular coverage to be divided into sectors, each with a separate feed, to obtain more nearly equal time delay for all beam positions. The feeds may be used to obtain simultaneous beams,⁵³ or one feed may be selected by a switching network.

Some applications of space-fed-array lenses are discussed in the literature.⁵⁴⁻⁵⁷

18.9 DOME ANTENNAS⁵⁸⁻⁶⁵

The dome antenna was developed to obtain an antenna to scan a beam over a full hemisphere or more for hemispherical-coverage applications as an alternative to the three or four planar arrays usually considered.

In its simplest form, the dome antenna (see Figure 18-19) has a constrained constant-thickness lens (the dome) fed by a planar array, both symmetric about the vertical axis. The technique is to illuminate an aperture (a portion of the dome) that has an appreciable area perpendicular to any of the desired beam directions to obtain antenna gain in those directions. The feed array can produce convergent or divergent rays to illuminate the required

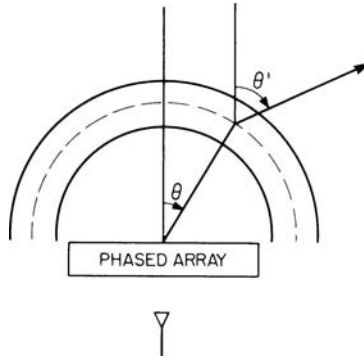


FIGURE 18-19 Dome-antenna nomenclature

lens area. The lens has a phase-delay gradient in the elevation plane to refract rays in the θ' direction when fed at an angle θ . $K = \theta'/\theta$ is defined as the scan amplification factor.

The dome is constructed of modules normal to the surfaces which have radiators on each surface connected by transmission lines that contain the required phase delay. The phase delay could be provided by switchable phase shifters,⁵⁸ but in practice this method is too expensive, so fixed delays are used. There are a number of methods for implementing the delays; in one module configuration,⁵⁹ the lengths of two dielectric-filled circular waveguides of different diameters in series are varied to obtain the phase delay in 20° increments. The modules may have phase delays modulo 360° (with a bandwidth penalty) and have different polarizations on the two lens surfaces. The spacing between modules must be small enough to suppress grating lobes.

Phase delays are generally equal in all modules that have the same cone angle θ . There are several design techniques for determining dome-module phase delays. Since the dome is in the near field of the feed array, ray-tracing methods may be used. For a hemispherical dome of radius R and a constant $K \neq 1$, one technique⁵⁸ equates the path lengths for rays from the center of the feed array at angles θ and $\theta + d\theta'$ and, after integrating, obtains the required phase delay, which is $(2\pi/\lambda) R [1 - \cos(\theta' - \theta)]/(K - 1)$. For these fixed phase delays, rays from the desired planar wavefront are traced through the lens to the feed array to determine the required feed-array wavefront, which in general is not planar. The gain (relative to the feed-array broadside gain without a dome) of an antenna designed in this manner is shown in Figure 18-20 for several values of K . It is evident that beam angles greater than 90° can be obtained for $K > 1$ and that maximum gain occurs for beam angles $\theta' > 90^\circ$ for $K > 2$.

It is desirable to be able to design for prescribed antenna gain versus θ' profiles as, for example,⁵⁸ that shown in Figure 18-21. For some applications, it is also helpful if the dome has an elongated shape along the axis; i.e., the dome is a cylinder generated by a noncircular arc. For these cases, the lens-module phase delays can be designed⁵⁸ by allowing a variable K with θ and applying the techniques used for designing a shaped-beam reflector (see Silver,² pages 497–500).

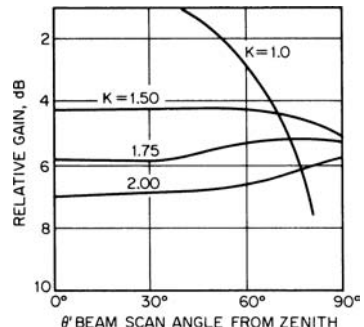


FIGURE 18-20 Dome-antenna relative gain versus θ' as a function of K

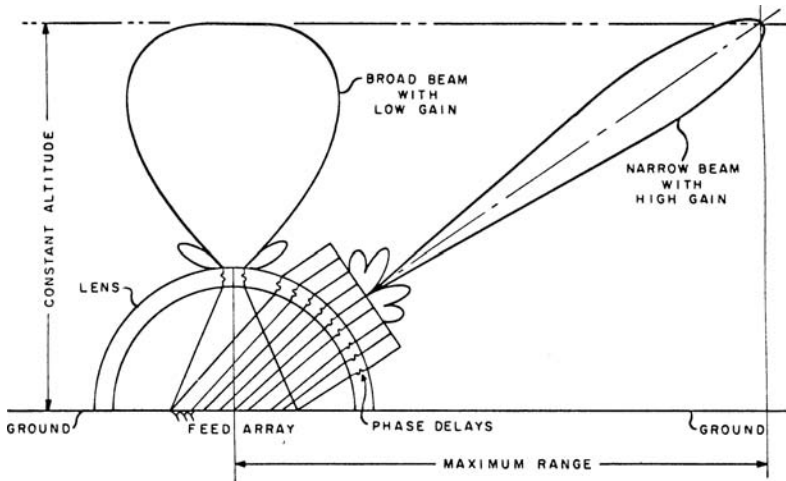


FIGURE 18-21 Typical gain versus θ' profile for a dome antenna

These techniques have been implemented⁵⁹ at C band with a 4-ft- (1.2-m-) diameter hemispherical dome composed of 3636 modules (dielectric-loaded circular waveguides with a phase-delay increment of 20°) and an $F/D = 0.75$, space-fed planar phased array with 805 elements. Dome loss was less than 1 dB over a 10 percent frequency band. Two gain-versus-scan-angle-profile designs were measured. One version with $K < 2$ with maximum gain at approximately 60° had peak sidelobes below -18 dB and average sidelobes below -25 dB. It had a 13 percent bandwidth for a zenith beam and a 6.3 percent bandwidth for a 60° beam position. The gain profile matched predictions within 0.3 dB for scan angles up to 60° . Another version with $K = 2.6$ with maximum gain at 110° demonstrated good performance in a limited amount of data. Another implementation,⁶⁰ also at C band, had a hemisphere on a cylinder dome 76 in (1.93 m) in diameter and 51 in (1.30 m) high that had 8511 modules, each a thin-walled metal tube that contained a serrated cruciform to provide the phase delay in increments of 20° between modules. The feed array was 47 in (1.19 m) in diameter and used 1120 elements. The gain profile had maximum gain at $\theta' = 90^\circ$.

An alternative to a constrained lens is the use of a graded-thickness (in elevation) homogeneous dielectric material with quarter-wavelength matching layers on each surface.^{58,65} Several techniques for designing to desired gain versus scan-angle profiles and for determining limits on attainable profiles have been analyzed.⁶¹⁻⁶³ A two-dimensional antenna with $K = 1.5$ that used a concave-outer-surface Archer lens for the feed has been analyzed⁶³ and measured⁶⁴; it improves the wide-scan-angle sidelobes, operates over a frequency band of more than an octave, and can provide multiple simultaneous beams.

18.10 LUNEBURG LENSES

Luneburg⁶⁶ investigated a class of spherically symmetric, variable-index-of-refraction n lenses that image two concentric spheres on each other. The *Luneburg lens* is a special case of this class in which one sphere has an infinite radius and the other is the lens surface. For a unit-radius lens, the n as a function of the radius r is $n = (2 - r^2)^{1/2}$. When fed at a point on the surface, it produces a plane wavefront, and by moving the feed on the surface, beam

scan over all space can be obtained (if the presence of the lens support and feed-scanning mechanism is ignored). The scanning applications and many design principles of Luneburg lenses are discussed in Chapter 18.⁶⁶

A number of authors^{67–75} have extended Luneburg's work. References 67–70 derive small-feed-circle lenses in which the feed radius is smaller than the lens radius. References 70–74 consider variations that include conical and beam-shaping wavefronts in addition to the plane wavefront. Cheng⁷⁵ considers a design to use a defocused feed.

Luneburg-lens radiation-pattern computations have been considered by most of the Luneburg-lens authors and in particular by Jasik⁷⁶ and Braun.⁷⁷ The Luneburg-lens aperture illumination is the feed pattern times $\sec \phi$, where ϕ is the feed-pattern angle. The increase in energy density near the aperture edges produces (relative to other lenses) high gain and small beamwidths and limits the sidelobe levels to -17 to -18 dB.

Early Luneburg lenses were two-dimensional (or cylindrical) slices through the center of the sphere because techniques were not available for constructing a sphere. One model⁷⁸ was a 36-in- (914-mm-) diameter lens constructed of almost parallel plates filled with polystyrene that operated in the TE_{10} mode at X band. The plate spacing was $a = \lambda/[2(\epsilon_r - 2 - r^2)^{1/2}]$, where ϵ_r is the dielectric constant of the filler material and r is the normalized radius. In addition to the aperture-edge energy increase, two-dimensional Luneburg lenses produce a saddle-shaped wavefront because of the circular aperture. This can be corrected by adding to the aperture parallel plates with a linear aperture, but this limits the beam-scan angle. Calculated and experimental radiation patterns show about -17 -dB sidelobes in the E plane and a flat-top pattern in the H plane. Jones⁷⁹ constructed an air-filled, almost parallel-plate TE_{10} -mode lens. Walter⁸⁰ constructed surface-wave lenses with tapered-thickness dielectric on a ground plane, metal posts on a ground plane, and a "holey-plate" structures.

The symmetry of the Luneburg lens can be exploited by adding planar reflectors through the lens center to obtain a virtual-source antenna.⁸¹ This permits eliminating large segments of the lens to reduce weight and allows the lens to be supported more easily, but it introduces aperture blocking by the feed.

Spherical Luneburg lenses have been constructed by using several techniques. Crushed Styrofoam⁸² was used for a 24-in- (610-mm-) diameter lens. Expandable polystyrene beads,^{83–85} expanded and solidified in heated hemispherical molds, were used for a 10-step, equal- ϵ_r increment, 18-in- (457-mm-) diameter lens. Webster⁸⁶ provides performance of a 10-step Styrofoam 18-in- (457-mm-) diameter lens. Foamed glass⁸⁷ was used to fabricate a 10-step 12-in- (305-mm-) diameter lens to obtain high-power operation. Calculations⁸⁸ of the heat distribution in a Luneburg lens indicate that the feed region reaches the highest temperature under high-power operation. A ray-tracing method for checking the design of a stepped- n Luneburg lens is provided in Mathis.⁸⁹

An important use of the Luneburg lens is as a reflector,^{85,90} obtained by placing a reflecting cap on the lens surface. An incoming plane wave is focused at a point on the cap and reflected; after transit through the lens the second time, it is a plane wave propagating in the opposite direction. The angular extent of the cap is chosen as a compromise between obtaining wide-angle performance and minimizing partial blockage of incoming rays by the cap.

For many applications that do not require the performance of a Luneburg lens, the literature^{91–95} investigates the performance of constant- n spherical lenses that are more easily constructed than Luneburg lenses.

Another class of two-dimensional Luneburg lenses is the geodesic lens, composed of a pair of nonplanar metal plates spaced a constant distance apart with a spacing small enough so that only a TEM mode can propagate. Myers⁴⁹ proved that rays in this medium follow geodesic paths on the mean surface. Rinehart^{96,97} derived the surface shape that is required for a two-dimensional Luneburg analog whose edge has a tangent normal to the aperture plane. Warren and Pinnell^{98,99} derived surfaces with the edge tangent parallel to the aperture plane.

Kunz¹⁰⁰ generalized the analysis and described several designs. Rudduck et al.¹⁰¹⁻¹⁰³ describe small-feed-circle design and also designs that radiate at a displaced angle from the aperture plane. Johnson¹⁰⁴ provides an excellent summary of geodesic lenses and distinguishes between tin-hat, helmet, and clamshell types. Hollis and Long¹⁰⁵ describe a feed-scanning implementation for limited beam-scan angles with a geodesic Luneburg lens. Johnson¹⁰⁶ shows radiation patterns of a 24-in- (610-mm-) diameter, $4.3\text{-mm-}\lambda$ lens with extension plates to a linear aperture that provides a total beam scan of 60° . The beamwidth is 0.5° , and sidelobes are below -25 dB for most beam positions. Johnson¹⁰⁷ demonstrates that any two-dimensional lens can have a geodesic analog.

18.11 FRESNEL ZONE PLATE LENSES

Lens antennas called *Fresnel zone plates* can be constructed in a planar form. They have the advantages of reduced weight, lower ohmic loss, and ease of construction compared with the conventional dielectric lenses shown in Figure 18-1. A Fresnel amplitude-zone plate lens consists of alternating opaque (metal) and transparent (dielectric) rings with the radii determined by the zone number. Measured results show that the beamwidth from these lenses is comparable with that obtained from a paraboloidal reflector of the same size but that the gain is about 15 dB less.¹⁰⁸ Fresnel phase-zone plate lenses are constructed by cutting annular rings in a flat piece of dielectric or by utilizing concentric rings of different dielectric materials of the same thickness.¹⁰⁹ Phase-zone plate lenses do not suffer the inefficiency of amplitude-zone plate lenses because the former are completely transparent.

REFERENCES

1. H. Jasik (ed.), *Antenna Engineering Handbook*, 1st Ed. (New York: McGraw-Hill, 1961). Chapter 14, "Lens-Type Radiators," by Seymour B. Cohn, is the basis for this chapter.
2. S. Silver, *Microwave Antenna Theory and Design*, MIT Rad. Lab. ser., vol. 12 (New York: McGraw-Hill, 1949). Chapter 11, by J. R. Risher, contains considerable basic information on dielectric, metal-plate, and waveguide lenses.
3. J. Brown, *Microwave Lenses* (London: Methuen & Co., Ltd., 1953).
4. R. C. Hansen, *Microwave Scanning Antennas*, vol. 1 (New York: Academic Press, Inc., 1964). Chapter 3, "Optical Scanners," by R. C. Johnson, includes material on Luneburg and bifocal dielectric lenses.
5. M. I. Skolnik, *Introduction to Radar Systems* (New York: McGraw-Hill, 1962). Chapter 7, "Antennas," has a section on lenses and many references.
6. M. I. Skolnik, *Radar Handbook* (New York: McGraw-Hill, 1970). Chapter 10, "Reflectors and Lenses," by D. C. Sengupta and R. E. Hiatt, discusses a number of lens types.
7. R. W. Kreutel, "The Hyperboloidal Lens with Laterally Displaced Dipole Feed," *IEEE Trans. Antennas Propagat.*, vol. AP-28 (July 1980): 443-450. Computes lens wavefront aberrations for off-axis feed positions.
8. W. E. Kock, "Metallic Delay Lenses," *Bell Syst. Tech. J.*, vol. 27 (January 1948): 58-82. This is the basic reference on artificial delay media.
9. F. G. Friedlander, "A Dielectric-Lens Aerial for Wide-Angle Beam Scanning," *J. IEE (London)*, part IIIA, vol. 93 (1946): 658. Discusses the design of dielectric lenses to satisfy the Abbe sine condition.
10. R. L. Sternberg, "Successive Approximation and Expansion Methods in the Numerical Design of Microwave Dielectric Lenses," *J. Math. & Phys.*, vol. 34 (January 1956): 209-235.

11. F. S. Holt and A. Mayer, "A Design Procedure for Dielectric Microwave Lenses of Large Aperture Ratio and Large Scanning Angle," *IRE Trans. Antennas Propagat.*, vol. AP-5 (January 1957): 25–30.
12. R. M. Brown, "Dielectric Bifocal Lenses," *IRE Conv. Rec.*, vol. 4, part 1 (1956): 180–187.
13. A. R. von Hippel, *Dielectric Materials and Applications* (New York: John Wiley & Sons, Inc., 1954).
14. E. M. T. Jones and S. B. Cohn, "Surface Matching of Dielectric Lenses," *J. App. Phys.*, vol. 26 (April 1955): 452–457.
15. T. Morita and S. B. Cohn, "Microwave Lens Matching by Simulated Quarter-Wave Transformers," *IRE Trans. Antennas Propagat.*, vol. AP-4 (January 1956): 33–39.
16. E. M. T. Jones, T. Morita, and S. B. Cohn, "Measured Performance of Matched Dielectric Lenses," *IRE Trans. Antennas Propagat.*, vol. AP-4 (January 1956): 31–33.
17. W. E. Kock, "Metal Lens Antennas," *IRE Proc.*, vol. 34 (November 1946): 828–836. This is the basic metal-plate-lens reference.
18. R. I. Primich, "A Semi-Infinite Array of Parallel Metallic Plates of Finite Thickness for Microwave Systems," *IRE Trans. Microwave Theory Tech.*, vol. MTT-4 (July 1956): 156–166. It is shown, for normal incidence and given n , that finite plate thickness can reduce reflections.
19. J. F. Carlson and A. E. Heins, "The Reflection of an Electromagnetic Plane Wave by an Infinite Set of Plates," *Q. App. Math.*, vol. 4 (1947): 313–329. Presents a mathematical solution for polarization perpendicular to the plane of incidence.
20. B. A. Lengyel, "Reflection and Transmission at the Surface of Metal-Plate Media," *J. App. Phys.*, vol. 22 (March 1951): 265–276. Extends Carlson and Hein's theory to take account of diffracted beams; presents theoretical formulas and experimental verification.
21. F. Berz, "Reflection and Refraction of Microwaves at a Set of Parallel Metallic Plates," *IEE Proc. (London)*, part III, vol. 98 (January 1951): 47–55. Presents analysis similar to that of Lengyel but with the restriction that the boundary surface is perpendicular to the plates.
22. E. A. N. Whitehead, "The Theory of Parallel-Plate Media for Microwave Lenses," *IEE Proc. (London)*, part III, vol. 98 (March 1951): 133–140. Presents analysis similar to that of Lengyel; contains many theoretical graphs for different angles of plate stagger and for typical lens shapes.
23. E. M. Wells, "Some Experiments on the Reflecting Properties of Metal-Tube Lens Medium," *Marconi Rev.*, vol. 17 (1954): 74–85. Gives experimental data for both polarizations for arrays of square tubes.
24. E. K. Proctor, "Methods of Reducing Chromatic Aberration in Metal-Plate Microwave Lenses," *IRE Trans. Antennas Propagat.*, vol. AP-6 (July 1958): 231–239. Slot-loaded ridge waveguides make possible greater bandwidth, improved scanning, and elimination of stepping.
25. R. L. Smedes, "High-Efficiency Metallized-Fiberglass Microwave Lens," *IRE Nat. Conv.*, New York, March 1956. A waveguide-type lens that utilizes dielectric and ridge loading, surface matching, and a variable refractive index is described.
26. H. E. King, J. L. Wong, R. B. Dybdal, and M. E. Schwartz, "Experimental Evaluation of a Circularly Polarized Metallic Lens Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-18 (May 1970): 412–414.
27. A. R. Dion, "An Investigation of a 110-Wavelength EHF Waveguide Lens," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (July 1972): 493–496.
28. J. Ruze, "Wide-Angle Metal-Plate Optics," *Ire Proc.*, vol. 28 (January 1950): 53–59. Gives design equations and experimental data for bifocal waveguide lenses designed for wide-scan-angle performance.
29. A. L. Peebles, "A Dielectric Bifocal Lens for MultiBeam Antenna Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-36, no. 5 (May 1988): 599–606.
30. C. M. Rappaport and A. I. Zaghoul, "Optimized Three-Dimensional Lenses for Wide-Angle Scanning," *IEEE Trans. Antennas Propagat.*, vol. AP-33, no. 11 (November 1985): 1227–1236.
31. E. K. Proctor and M. H. Rees, "Scanning Lens Design for Minimum Mean-Square Phase Error," *IRE Trans. Antennas Propagat.*, vol. AP-5 (October 1957): 348–355. A calculus-of-variations analysis leads to design parameters close to those of Ruze.

32. E. Fine and G. Reynolds, "A Point Source Bi-Normal Lens," Air Force Cambridge Res. Cen. Rep. E-5095, Bedford, Mass., May 1953.
33. G. D. M. Peeler and W. F. Gabriel, "Volumetric Scanning GCA Antenna," *IRE Conv. Rec.*, vol. 3, part 1 (1955): 20–27.
34. A. R. Dion and L. J. Ricardi, "A Variable-Coverage Satellite Antenna System," *IEEE Proc.*, vol. 59 (February 1971): 252–262.
35. H. S. Lu, "On Computation of the Radiation Pattern of a Zoned Waveguide Lens," *IEEE Trans Antennas Propagat.*, vol. AP-22 (May 1974): 483–484.
36. C. F. Winter, "A TE-Mode Parallel Plate Lens," USAF Antenna Symp., University of Illinois, Monticello, October 1972.
37. J. S. Ajioka and V. W. Ramsey, "An Equal Group Delay Waveguide Lens," *IEEE Trans Antennas Propagat.*, vol. AP-26 (July 1978): 519–527.
38. A. R. Dion, "A Broadband Compound Waveguide Lens," *IEEE Trans Antennas Propagat.*, vol. AP-26 (September 1978): 751–755.
39. H. Gent, "The Bootlace Aerial," *Roy. Radar Establishment J.*, Malvern, England, no. 40 (October 1957): 47–58.
40. W. Rotman and R. F. Turner, "Wide-Angle Microwave Lens for Line Source Applications," *IEEE Trans Antennas Propagat.*, vol. AP-11 (November 1963): 623–632. Derives and discusses the Rotman-Turner lens.
41. D. Archer, "Lens-Fed Multiple-Beam Arrays," *Electron. Prog. (Raytheon Co.)* (winter 1974): 24–32; and *Microwave J.* (October 1975): 37–42. Describes the Archer, R-2R, and R-KR lenses and methods of stacking lenses to obtain beam clusters.
42. D. H. Archer, R. J. Prickett, and C. P. Hartwig, Multi-Beam Array Antenna, (September 25, 1973): U.S. Patent 3,761,936. This is the patent for the Archer lens.
43. D. T. Thomas, "Multiple Beam Synthesis of Low Sidelobe Patterns in Lens Fed Arrays," *IEEE Trans Antennas Propagat.*, vol. AP-26 (November 1978): 883–886. Describes techniques for feeding several beam ports in an Archer lens to obtain low sidelobe radiation patterns.
44. J. B. L. Rao, "Multifocal Three Dimensional Bootlace Lenses," *IEEE PGAP Nat. Conv. Rec.* (1979): 332–335. Provides design equations for two-, three-, and four-focal-point bootlace lenses.
45. J. P. Shelton, "Focusing Characteristics of Symmetrically Configured Bootlace Lenses," *IEEE Trans Antennas Propagat.*, vol. AP-26 (July 1978): 513–518. Derives design equations for a new type of antenna that has two (symmetrical) bootlace lenses, one on the feed-port side and one for the aperture.
46. J. P. Shelton, "Three-Dimensional Bootlace Lenses," *IEEE PGAP Nat. Conv. Rec.* (1980): 568–571. Provides design equations for several three-dimensional bootlace lenses.
47. W. Rotman and P. Franchi, "Cylindrical Microwave Lens Antenna for Wideband Scanning Applications," *IEEE PGAP Nat. Conv. Rec.* (1980): 564–567. Provides designs for a bootlace-type lens with phase shifters and a limited number of feeds to provide coverage over wide-scan angles without requiring a feed for each beam position.
48. H. B. Devore and H. Iams, "Microwave Optics between Parallel Conducting Sheets," *RCA Rev.*, vol. 9 (December 1948): 721–732. Discusses several microwave optics devices including the R-2R geodesic scanner.
49. S. B. Myers, "Parallel Plate Optics for Rapid Scanning," *J. App. Phys.*, vol. 18, no. 2 (February 1947): 211. Discusses the general figure-of-revolution scanner formed as the mean surface of a parallel-plate region. This article discusses the scanner of Devore and Iams and considers other scanners in which the mean surface is a cone.
50. M. L. Kales and R. M. Brown, "Design Consideration for Two-Dimensional Symmetric Bootlace Lenses," *IEEE Trans Antennas Propagat.*, vol. AP-13 (July 1965): 521–528. Proves that the R-2R lens of Gent is unique.
51. P. J. Kahrilas, "HAPDAR—An Operational Phased Array Radar," *IEEE Proc.*, vol. 56 (November 1968): 1967–1975.
52. B. R. Hatcher, "Collimation of Row-and-Column Steered Phased Arrays," *IEEE Proc.*, vol. 56 (November 1968): 1787–1790.

53. R. Tang, E. E. Barber, and N. S. Wong, "A Wide Instantaneous Bandwidth Space Fed Antenna," *Eascon Rec.* (1975): 104A–104F. The Butler matrix provides multiple feeds for equal time delays.
54. E. J. Daly and F. Steudel, "Modern Electronically-Scanned Array Antennas," *Electron. Prog. (Raytheon Co.)* (Winter 1974): 11–17.
55. D. K. Barton, "Radar Technology for the 1980s," *Microwave J.* (November 1978): 81–86.
56. T. E. Walsh, "Military Radar Systems: History, Current Position, and Future Forecast," *Microwave J.* (November 1978): 87–95.
57. D. K. Barton, "Historical Perspective on Radar," *Microwave J.* (August 1980): 21–38.
58. J. J. Stangel and P. A. Valentino, "Phased Array Fed Lens Antenna," (August 28, 1973): U.S. Patent 3,755,815. This is the basic patent on the dome antenna.
59. L. Schwartzman and J. Stangel, "The Dome Antenna," *Microwave J.* (October 1975): 31–34. Describes and gives experimental data on a particular dome-antenna design.
60. L. Schwartzman and P. M. Liebman, "A Report on the Sperry Dome Radar," *Microwave J.* (March 1979): 65–69. Presents experimental data on one design of the dome antenna integrated into a radar.
61. L. Susman and H. Mieras, "Results of an Exact Dome Antenna Synthesis Procedure," *IEEE PGAP Nat. Symp. Rec.* (1979): 38–41. Presents a synthesis of gain versus scan angle for dome antennas.
62. H. Steyskal, A. Hessel, and J. Shmoys, "On the Gain-versus-Scan Tradeoffs and the Phase Gradient Synthesis for a Cylindrical Dome Antenna," *IEEE Trans Antennas Propagat.*, vol. AP-27 (November 1979): 825–831.
63. D. T. Thomas, "Design Studies of Wide Angle Array Fed Lens," *IEEE PGAP Nat. Symp. Rec.* (1979): 340–343.
64. D. T. Thomas and S. D. Bixler, "Hardware Demonstration of a 2-Dimensional Wide Angle Array Fed Lens," *IEEE PGAP Nat. Symp. Rec.* (1979): 344–347.
65. P. A. Valentino, C. Rothenberg, and J. J. Stangel, "Design and Fabrication of Homogeneous Dielectric Lenses for Dome Antennas," *IEEE PGAP Nat. Symp. Rec.* (1980): 580–583. Discusses a nonconstrained dome lens of graded-thickness dielectric materials.
66. R. K. Luneburg, *Mathematical Theory of Optics* (Berkeley: University of California Press, 1964); and mimeographed lecturer notes (Providence, R.I.: Brown University Press, 1944). This is the basic reference on the Luneburg lens.
67. J. Brown, "Microwave Wide Angle Scanner," *Wireless Eng.*, vol. 30, no. 10 (October 1953): 250–255. Discusses the small-feed-circle Luneburg lens.
68. A. S. Gutman, "Modified Luneburg Lens," *J. Appl. Phys.*, vol. 25 (1954). Discusses the small-feed-circle Luneburg lens.
69. A. F. Kay, "The Impossibility of Certain Desirable Luneburg Lens Modifications," *IRE Trans Antennas Propagat.*, vol. AP-4 (January 1956): 87–88.
70. J. E. Eaton, "An Extension of the Luneburg-Type Lenses," Nav. Res. Lab. Rep. 4110, 1953. Discusses the small-feed-circle Luneburg lens and lenses that produce wavefronts other than the plane wave.
71. S. P. Morgan, "General Solution of the Luneburg Lens Problem," *J. Appl. Phys.*, vol. 29 (September 1958): 1358–1368.
72. J. R. Huynen, "Theory and Design of a Class of Luneburg Lenses," *IRE Wescon Conv. Rec.*, vol. 2, part 1 (1958): 219–230.
73. A. F. Kay, "Spherically Symmetric Lens," *IRE Trans Antennas Propagat.*, vol. AP-7 (January 1959): 32–38.
74. S. P. Morgan, "Generalizations of Spherically Symmetric Lenses," *IRE Trans Antennas Propagat.*, vol. AP-7 (October 1959): 342–345.
75. D. H. Cheng, "Modified Luneburg Lens for Defocused Source," *IRE Trans Antennas Propagat.*, vol. AP-8 (January 1960): 110–111.
76. H. Jasik, "The Electromagnetic Theory of the Luneburg Lens," Air Force Cambridge Res. Cen. Rep. TR-54-121, Bedford, Mass., November 1954.

77. E. H. Braun, "Radiation Characteristics of the Spherical Luneburg Lens," *IRE Trans Antennas Propagat.*, vol. AP-4 (April 1956): 132–138.
78. G. D. M. Peeler and D. H. Archer, "A Two-Dimensional Microwave Luneburg Lens," *IRE Trans Antennas Propagat.*, vol. AP-1 (July 1953): 12–23. Discusses a two-dimensional Luneburg lens constructed of almost parallel plates filled with dielectric material.
79. S. S. D. Jones, "A Wide Angle Microwave Radiator," *IEE Proc. (London)*, part III, vol. 97 (1950): 225.
80. C. H. Walter, "Surface-Wave Luneburg Lens Antennas," *IRE Trans Antennas Propagat.*, vol. AP-8 (September 1960): 508–515.
81. G. D. M. Peeler, K. S. Kelleher, and H. P. Coleman, "Virtual Source Luneburg Lenses," *IRE Trans Antennas Propagat.*, vol. AP-2 (July 1954): 94–99. Discusses the Luneburg lens with metallic planes through the center.
82. G. P. Robinson, "Three-Dimensional Microwave Lens," *Tele-Tech & Electronic Ind.* (November 1954): 73.
83. G. D. M. Peeler and H. P. Coleman, "Microwave Stepped-Index Luneburg Lenses," *IRE Trans Antennas Propagat.*, vol. AP-6 (April 1958): 202–207.
84. M. C. Volk and G. D. M. Peeler, "A Three-Dimensional Microwave Luneburg Lens," URSI meeting, Washington, 1955. Discusses a lens constructed of hemispherical shells of increasing radius and decreasing dielectric constant.
85. E. F. Buckley, "Stepped-Index Luneburg Lenses," *Electron. Design*, vol. 8 (Apr. 13, 1960): 86–89.
86. R. E. Webster, "Radiation Patterns of a Spherical Luneburg Lens with Simple Feeds," *IRE Trans Antennas Propagat.*, vol. AP-6 (July 1958): 301–302.
87. L. C. Gunderson and J. F. Kauffman, "A High Temperature Luneburg Lens," *IEEE Proc.* (May 1968): 883–884.
88. D. S. Lerner, "Calculation of Radiation Heating in a Microwave Luneburg Lens," *IEEE Trans Antennas Propagat.*, vol. AP-12 (January 1964): 16–22.
89. H. F. Mathis, "Checking Design of Stepped Luneburg Lens," *IRE Trans Antennas Propagat.*, vol. AP-8 (May 1960): 342–343.
90. J. I. Bohnert and H. P. Coleman, "Applications of the Luneburg Lens," Nav. Res. Lab. Rep. 4888, Mar. 7, 1957.
91. G. Bekefi and G. W. Farnell, "A Homogeneous Dielectric Sphere as a Microwave Lens," *Can. J. Phys.*, vol. 34 (August 1956): 790–803.
92. S. Cornbleet, "A Simple Spherical Lens with External Foci," *Microwave J.* (May 1965): 65.
93. T. L. ap Rhys, "The Design of Radially Symmetric Lenses," *IEEE Trans Antennas Propagat.*, vol. AP-18 (July 1970): 497–506.
94. L. C. Gunderson, "An Electromagnetic Analysis of a Cylindrical Homogeneous Lens," *IRE Trans Antennas Propagat.*, vol. AP-20 (July 1972): 476–479.
95. W. R. Free, F. L. Cain, C. E. Ryan, Jr., C. P. Burns, and E. M. Turner, "High-Power Constant-Index Lens Antennas," *IEEE Trans Antennas Propagat.*, vol. AP-22 (July 1974): 582–584.
96. R. F. Rinehart, "A Solution of the Rapid Scanning Problem for Radar Antennae," *J. App. Phys.*, vol. 19 (September 1948): 860–862. Derives a mean surface in a parallel-plate region that is an analog of a two-dimensional Luneburg lens.
97. R. F. Rinehart, "A Family of Designs for Rapid Scanning Radar Antennas," *IRE Proc.*, vol. 40, no. 6 (June 1952): 686–687. Generalizes the Luneburg analog for a small feed circle.
98. F. G. R. Warren and S. E. A. Pinnell, "Tin Hat Scanning Antennas," Tech. Rep. 6, RCA Victor Co., Ltd., Montreal, 1951. Presents an analog of the Luneburg lens that has tangents at its periphery in the lens-aperture plane.
99. F. G. R. Warren and S. E. A. Pinnell, "The Mathematics of the Tin Hat Scanning Antenna," Tech. Rep. 7, RCA Victor Co., Ltd., Montreal, 1951.
100. K. S. Kunz, "Propagation of Microwaves between a Parallel Pair of Doubly Curved Conducting Surfaces," *J. Appl. Phys.*, vol. 25 (May 1954): 642–653. Discusses geodesic path-length lenses equivalent to the Luneburg lens.

101. R. C. Rudduck and C. H. Walter, "A General Analysis of Geodesic Luneburg Lenses," *IRE Trans Antennas Propagat.*, vol. AP-10 (July 1962): 444–450.
102. R. C. Rudduck, C. E. Ryan, Jr., and C. H. Walter, "Beam Elevation Positioning in Geodesic Lenses," *IEEE Trans Antennas Propagat.*, vol. AP-12 (November 1964): 678–684.
103. G. A. Thiele and R. C. Rudduck, "Geodesic Lens Antennas for Low-Angle Radiation," *IEEE Trans Antennas Propagat.*, vol. AP-13 (July 1965): 514–521.
104. R. C. Johnson, "The Geodesic Luneburg Lens," *Microwave J.* (August 1962): 76–85.
105. J. S. Hollis and M. W. Long, "A Luneburg Lens Scanning System," *IRE Trans Antennas Propagat.*, vol. AP-5 (January 1957): 21–25.
106. R. C. Johnson, "Radiation Patterns from a Geodesic Luneburg Lens," *Microwave J.* (July 1963): 68–70.
107. R. C. Johnson and R. M. Goodman, Jr., "Geodesic Lenses for Radar Antennas," *Eascon Rec.* (1968): 64–69.
108. L. F. Van Buskirk and C. E. Hendrix, "The Zone Plate as a Radio-Frequency Focusing Element," *IRE Trans Antennas Propagat.*, vol. AP-9 (May 1961): 319–320.
109. D. N. Black and J. C. Wiltse, "Millimeter-Wave Characteristics of Phase-Correcting Fresnel Zone Plates," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35 (December 1987): 1122–1129.

Chapter 19

Ultrawide Bandwidth
Antenna Design

Chi-Chih Chen
The Ohio State University

CONTENTS

19.1	INTRODUCTION.....	19-2
19.2	FUNDAMENTALS OF UWB ANTENNA DESIGN.....	19-2
19.3	TYPE-I UWB ANTENNA DESIGN	19-5
19.4	TYPE-II UWB ANTENNA DESIGN	19-6
19.5	TYPE-III UWB ANTENNA DESIGN	19-7
19.6	THE DUAL-POLARIZATION UWB DIELECTRIC ROD PROBE (DRP) ANTENNA	19-9
19.7	THE DUAL-POLARIZATION UWB DIELECTRIC HORN ANTENNA DESIGN	19-13

19.1 INTRODUCTION

There has been continuing interest in designing a UWB antenna that operates over a wide frequency range and that can be used for multiple channels or systems. Regulatory restrictions on bandwidth and the greater technical challenges and cost involved in developing a UWB antenna, as well as the corresponding operating systems, have largely limited the development of UWB antennas to special research organizations and agencies in military applications. A recent FCC ruling allowing the operation of UWB signals in various sensing, imaging, and communication applications has once again revived interest in designing UWB antennas and in their potential applications. Technology advances in high-speed analog-to-digital and digital-to-analog devices, digital frequency synthesis, and digital receivers also allow for rapid and low-cost implementation of UWB systems below 3 GHz. Therefore, more attention is now being given to designing compact and low-cost UWB antennas.

The term *UWB* has often been used loosely. The original FCC definition of UWB specifies 25 percent or more fractional bandwidth. That is the ratio of a signal's -10 -dB bandwidth to center frequency. In this chapter, *UWB* mainly refers to a bandwidth of more than 2:1. That is the ratio of the highest frequency to the lowest frequency of the -10 -dB bandwidth. Some people determine the antenna's bandwidth based on the frequency range where the return loss (or reflection coefficient) level is less than -10 dB. However, note that a low return loss does not necessarily imply a strong radiation. Nor does it say anything about the radiation pattern. The return loss value has even less meaning if the antenna contains absorptive mechanisms. Since the main function of an antenna is to radiate sufficient energy in desired directions, it makes sense to use realized gain and radiation pattern to define the bandwidth.

There are many theories and examples of UWB antenna designs. The literature also offers many detailed discussions about design procedures and the parameters of these designs. Unfortunately, many of them fall short of providing physical insights about antenna operations and wave behaviors. Modern powerful computation capability seems to turn more antenna designers into antenna optimizers who seek to rely on massive computation power to come up with an "improved" or "optimized" antenna design. No doubt, this approach is indeed attractive to young antenna engineers who have not yet developed a good physical understanding of antennas, especially for complex antennas, but who have modeling tools and experience. Therefore, it is not the intention of this chapter to discuss various specific UWB antenna designs; there are already many great discussions in other chapters of this Handbook. Rather, this chapter attempts to simplify discussions by focusing on the fundamental principles and formations of UWB antennas to promote a better physical understanding of UWB antenna operations and to encourage the development of more new UWB antenna designs.

19.2 FUNDAMENTALS OF UWB ANTENNA DESIGN

Before we discuss the principles of designing a UWB antenna, it is essential to review the phenomena of propagation, scattering, and radiation of electromagnetic waves since they will provide important clues to UWB antenna operations.

What Is Propagation of Electromagnetic Waves?

The propagation of electromagnetic waves is an iterative induction process of time-varying electric and magnetic fields, which is governed by the following equations in a source-free region:

$$\begin{aligned}\nabla \times \bar{E} &= -\frac{\partial(\mu \bar{H})}{\partial t} \\ \nabla \times \bar{H} &= \frac{\partial(\epsilon \bar{E})}{\partial t}\end{aligned}\quad (19-1)$$

For instance, Figure 19-1 illustrates that a time-varying electric field in a source-free region induces time-varying magnetic fields around it. These induced magnetic fields subsequently induce time-varying electric fields around them. This induction process continues and constitutes the electromagnetic propagation. Any change in the property of the medium where propagation is taking place causes electromagnetic scattering. Such scattering could be in the form of reflection, refraction, or diffraction and depends upon the type of change involved, as illustrated in Figure 19-2. It is helpful to understand such scattering behaviors while designing a UWB antenna since an antenna can be considered as a scattering structure strategically positioned near an excitation source. The total radiated field is then the sum of the unperturbed incident field \bar{E}^i and the scattered field \bar{E}^s , that is, $\bar{E} = \bar{E}^i + \bar{E}^s$.

The properties (spectrum, pattern, amplitude, and phase) of the scattered fields depend on the geometry, composition, and orientation of the perturbation with respect to the incident fields.

What Is Electromagnetic Radiation?

Electromagnetic radiation is a phenomenon of unbounded propagation of electromagnetic waves. Radiation causes a net loss of electromagnetic energy from the source region (a region where electromagnetic energy is supplied). The total radiated power is calculated from total power flow away from an arbitrarily large closed region that contains sources, that is, $P_r = \oint_S \bar{E} \times \bar{H}^* \cdot d\bar{S}$.

What Is an Antenna?

There are many ways to define an antenna. Some consider an antenna as an impedance transformer between a transmission line and the open space. Others define an antenna as a coupler that transfers electromagnetic energy into space. Readers who use most antennas

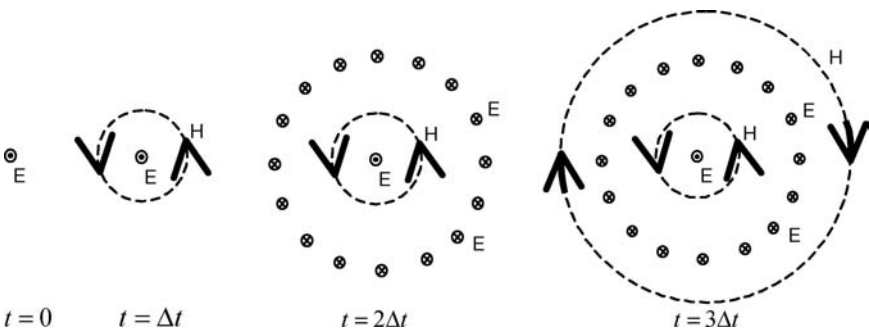


FIGURE 19-1 The iterative induction process of electromagnetic propagations

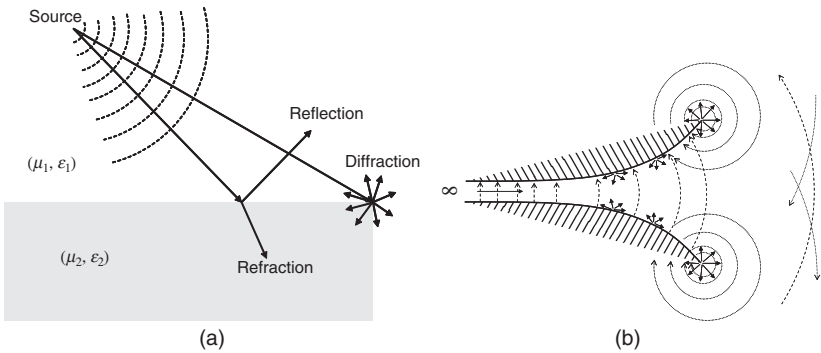


FIGURE 19-2 Scattering of electromagnetic waves: (a) material changes and (b) geometry changes

made of good (or poor) conductors probably are used to thinking of an antenna as a current-guiding structure that somehow converts the conduction currents to displaced currents. It is probably better to consider an antenna as a single or plural near-field scattering structure that generates radiation by the strategic scattering of the incident fields produced by an excitation source. Under such considerations it is obvious that an antenna does not have to be made of conductors. Instead any materials that have the right properties for creating and controlling scattering can be used to design an antenna. Note that this is different from using materials as an additional loading. Rather, the material itself could be the primary component of the antenna structure. For instance, two UWB antenna examples discussed at the end of this chapter utilize a low-loss dielectric body as the main radiating structures. This understanding is important since many advanced current or future materials have unique properties that could be utilized for antenna designs.

It is also important to recognize that the difference between a transmission line and an antenna lies only in the amount of radiation generated, not in the geometry itself! For instance, an open-ended, two-wire “transmission line” can be an “antenna” if the spacing between the lines becomes large with respect to the wavelength such that the diffraction from the wires and wire ends do not cancel each other or the incident fields in the exterior region. For the same reason, a “horn antenna” may become just an open-ended “waveguide” if the aperture becomes small compared with the wavelength.

What Makes a UWB Antenna?

A UWB antenna is an antenna capable of producing a *similar radiation* (patterns and gain) over a very wide frequency range (more than 2:1 bandwidth). Although a wider bandwidth can be obtained by trading off impedance matching and realized gain, we will not consider this approach in this chapter. We will refer to the *realized gain*, $G = e(1-|\Gamma|^2)D$, as simply *gain* throughout this chapter. The variables e , Γ , and D represent antenna efficiency, input reflection coefficient (or return loss), and directivity, respectively. It should be cautioned that defining an antenna’s bandwidth based on return loss being lower than -10 dB is inadequate since it does not guarantee strong radiation in the desired direction(s). When an antenna contains absorptive materials, low return loss indicates low return and says nothing about radiation (gain, bandwidth, and pattern). UWB radiation can be accomplished via the following three approaches:

- I. Perturbing electromagnetic resonance to broaden the resonant peak
- II. Allowing only one dominant radiation region that is physically small compared with wavelengths
- III. Maintaining a “similar” radiation/scattering geometry (shape and dimension) in terms of wavelength

Herein, UWB antenna designs using the preceding *I*, *II*, or *III* approach will be referred to as Type-I, Type-II, and Type-III UWB antennas, respectively. It is unusual for Type-I UWB antennas to have more than 3:1 bandwidth. Type-II UWB antennas offer superior pulse response but may have low efficiency (if absorptive loading is applied) or frequency-dependent gain and patterns. Many Type-II antennas can achieve 9:1 bandwidth. Type-III UWB antennas offer frequency-independent gain and patterns but are often dispersive. The following sections will provide more analysis of these three types of UWB antennas.

19.3 TYPE-I UWB ANTENNA DESIGN

This type of UWB antenna achieves bandwidth by introducing incoherent resonance and effectively lowering the quality factor of the electromagnetic resonance. The majority of Type-I UWB antennas are variations of dipoles or monopoles.

Common Broadband Dipoles

The radiation of a thin-wire dipole is mainly from the superposition of the direct radiation (the incident field) from the feed and strong diffractions from the two ends. The magnitude and phase relationship between these diffracted fields with respect to each other and the direct incident field ultimately determine its bandwidth and pattern behavior, as illustrated in Figure 19-3a. The phase relationship is determined from the propagation delay from the feed to each end and then to a certain pattern direction. Notice that if the length, l , is short such that its propagation phase delay is insignificant, and if these diffractions are strong, they could jointly cancel the incident fields and result in weak radiation. This is indeed the case when the dipole is much less than half of a wavelength. Along the boresight, the interference between the incident and scattered fields is responsible for the narrowband and strong frequency-dependent gain and pattern behavior of the thin-wire dipole. It is well known that a bowtie dipole design has a much broader bandwidth compared with a

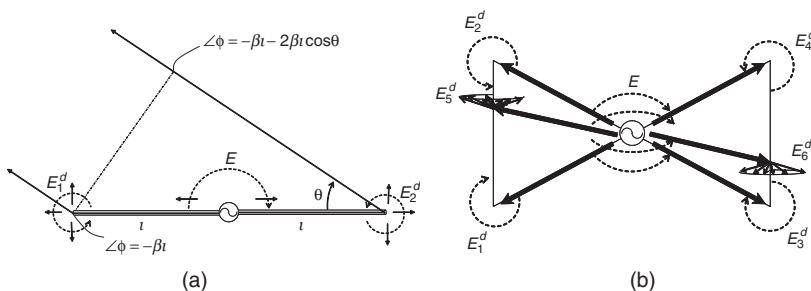


FIGURE 19-3 Comparison of the radiation/scattering of a thin-wire dipole and a bowtie dipole

thin-wire dipole. This is because of the involvement of more incoherent diffractions, as illustrated in Figure 19-3*b*, which shows four corner diffraction terms and two edge diffraction terms. Since not all diffractions have the same phase at a specific frequency and specific pattern angle, the total radiated fields exhibit a wider bandwidth. In fact, the wider the phase distribution of all contributing scattering terms, the wider the bandwidth. There are many other published 2D or 3D variations of dipoles¹⁻³ that showed bandwidths of up to 3:1. The upper frequency limit of this type of UWB antenna is usually limited by the excitation of higher modes that cause pattern narrowing and multiple radiation lobes.

19.4 TYPE-II UWB ANTENNA DESIGN

Type-II UWB antennas achieve wide bandwidth by controlling diffractions on antenna structures via careful geometry design or proper utilization of absorptive loading. Type-II UWB antennas often produce some degree of frequency-dependent gain and patterns because diffraction and absorption are functions of frequency.⁴ For instance, a curved section of an antenna may appear to be smooth for short wavelengths and produce less diffraction but would abruptly bend for long wavelengths. Similarly, a section of tapered resistive loading could be a long and smooth tapering for short wavelengths but would be short and abrupt for long wavelengths.

Type-II UWB Antenna Using Absorptive Loading

Instead of trying to add multiple diffractions to achieve a wide bandwidth as in Type-I UWB antennas, Type-II UWB antennas aim to minimize diffractions. Absorptive loading is an effective way to achieve this goal at the price of lower antenna efficiency. Therefore, the art of absorptive loading is to utilize the minimum amount of absorptive treatment to reduce diffractions to acceptable levels. This would require some knowledge of the types and causes of dominant diffraction events. Both electric and magnetic absorptive materials have been used. Resistively loaded dipoles⁵⁻⁸ and resistively loaded horns⁹ are well-known examples. Although the low efficiency can sometimes be compensated for by means of higher input power, caution needs to be exercised to take care of the heat produced by the absorptive loading.

Type-II UWB Antenna Using Curved Arms

Another way to reduce the diffraction effect without utilizing absorptive loading is to control diffractions by properly shaping antenna geometry to avoid strong localized diffractions, which usually arise from abrupt discontinuities. Instead, gentle and continuous diffractions are allowed to take place along the whole curved arms (see Figure 19-4). The low-frequency limit of this design occurs when the arm length is less than half of the wavelength. With such a short length, the curvature becomes ineffective, and significant diffractions arise at the ends of antenna arms. The high-frequency limit is determined by the pattern distortions caused

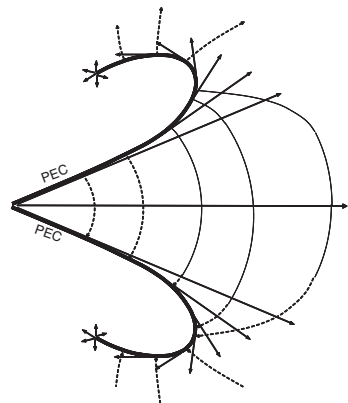


FIGURE 19-4 Smooth diffraction control using curved arms

by the curvature smoothness and the starting curvature. A properly curved antenna arm could increase the -10 -dB return-loss bandwidth to 9:1. However, such a design often exhibits frequency-dependent gain and pattern. Typically, the resultant boresight gain increases with frequency and beamwidth decreases with frequency. Common examples of Type-II UWB antennas include tapered-slot antennas,^{10–12} rolled-edge horns,^{13,14} and ridged horns.^{15–17} Of course, both curved-arm and absorptive loading can also be applied simultaneously to further reduce antenna size and to achieve a better pattern stability at the expense of antenna efficiency.⁹

19.5 TYPE-III UWB ANTENNA DESIGN

Type-III UWB antennas are probably the most widely used UWB antennas. This type of antenna adopts frequency-independent or frequency-scaled geometries. Among them, angle-defined geometry, complementary geometry, and log-periodic geometry^{18–20} are the best known. Although these geometries could be good candidates for designing a frequency-independent or UWB antenna, employing these geometries alone does not necessarily produce a UWB antenna design. Other antenna design considerations such as feeding scheme, mutual coupling, and truncations must be taken into account. This section discusses some basic guidelines that should be helpful in designing or optimizing a Type-III UWB antenna, which is characterized by its frequency-independent radiation characteristics such as gain, patterns, and impedance.

- **Frequency-Scaled Radiation Structure** Considering the far fields radiated from a local one-dimensional distribution of electromagnetic fields of equivalent currents:

$$\vec{H} = \nabla \times \left(\hat{x} \frac{e^{-jkr}}{4\pi r} \int_{-a_e}^{a_e} J(k_x, x') e^{jk_x x'} dx' \right) = D \frac{e^{-jkr}}{4\pi r} F(k_x a_e) \quad (19-2)$$

$$2a_e : \text{effective aperture size}; \quad k_x = k \sin \theta = \frac{2\pi}{\lambda} \sin \theta$$

the effective aperture area is the region where most in-phase radiations (at a given frequency) occur. From Eq. 19-2, if $J(k_x, x') \approx J(k_x x')$ and $a/\lambda \approx \text{constant}$, the radiated fields should be independent of frequency. That is, as frequency varies, the equivalent current distribution should be scaled in proportion to the wavelength.

- **Effective Radiation from Active Regions** At a given frequency, the antenna should form a region where all radiations are in phase and in the same polarization at any instant. Note that the phase and polarization of the radiated field could vary with time as in the case of circularly polarized antennas. Such a region is commonly called an *active region*. However, this terminology is somewhat misleading since radiation also occurs outside the active region. But the contribution of radiation from outside the active region is negligible in the far field because of cancellation or because it is too weak. Figure 19-5a shows a log-periodic dipole array (LPDA) and the active region where the currents on adjacent dipole elements become in phase at a given frequency and at any time. The location of the active region is related to the phase variation along the feeding transmission line also illustrated in Figure 19-5a. To increase radiation, the elements in the active region should also be at resonance. A strong radiation from the active region is very important in designing a Type-III UWB antenna because it ensures rapid energy attenuation after passing through the active region. Otherwise, frequency-dependent gain,

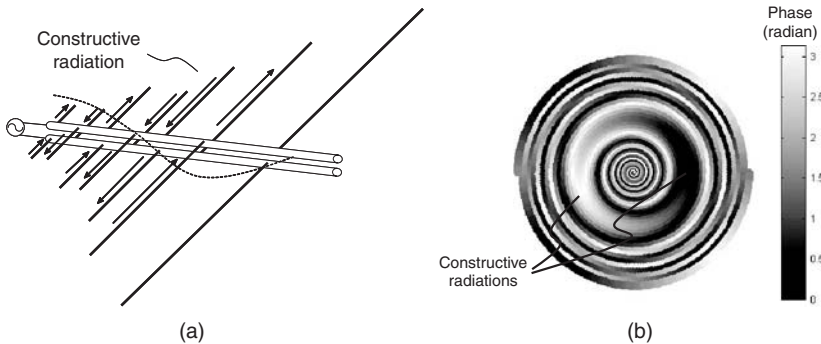


FIGURE 19-5 (a) Illustration of a time snapshot current direction and feeding-line phase on a log-periodic dipole array at a given frequency; (b) A time snapshot of the phase associated with fields on the arms of a circular Archimedean spiral antenna at a given frequency

pattern, and impedance could occur due to radiation from higher-order modes or from the ends of the antenna structure. Similarly, Figure 19-5b plots the instantaneous phase of the fields along a circular Archimedean spiral antenna at a given frequency. A broad active region can be clearly identified by its relatively slow phase variation. Note that although the left and right sides of the active region appear to be out of phase, the radiated fields are actually in phase since the current flows in opposite directions. Therefore, this active region radiates linearly polarized fields similar to a one-wavelength circular ring at any given time. As time progresses, the fields in the active region rotate and produce circularly polarized radiations.

- Strong Coupling Between Adjacent Modes** Continuous and stable antenna characteristics across an ultrawide bandwidth require a smooth transition from one active region to the next as frequency varies. This implies a strong coupling between adjacent structures. In the case of a spiral antenna (see Figure 19-5b), this requires a sufficient winding tightness (or lower growth rate) to avoid under sampling the phase variations in the active region. Such an under sampling causes undesired fluctuations in gain, pattern, and impedance.
- Excite from Highest Frequency End** Since an electrically large structure could support undesired higher modes, the antenna should be fed from the high-frequency region (small resonant structures). This region will be too small to support a resonance, so most of the lower-frequency energy will simply pass through to the larger structures.
- Active Region Transition Scheme** Consider a UWB antenna that is formed by assembling structures with similar geometry but which are scaled with distance from a reference feedpoint, as illustrated in Figure 19-6. Now, let's think about how the active region should migrate with frequency. Figure 19-7 reveals two possible schemes, *A* and *B*, with the shadowed region being the active region. At the highest frequency (leftmost figures), the active region is located near the center, where the feed is. As frequency decreases (middle figures), the active region in scheme *B* expands and encompasses the previous active region at a higher frequency. Although the "area" of the active region in the *B* scheme appears to be scaled with

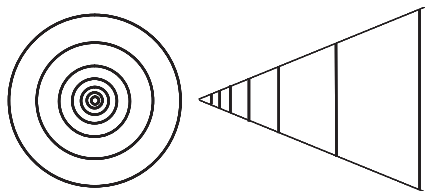


FIGURE 19-6 Distance-scaled structures

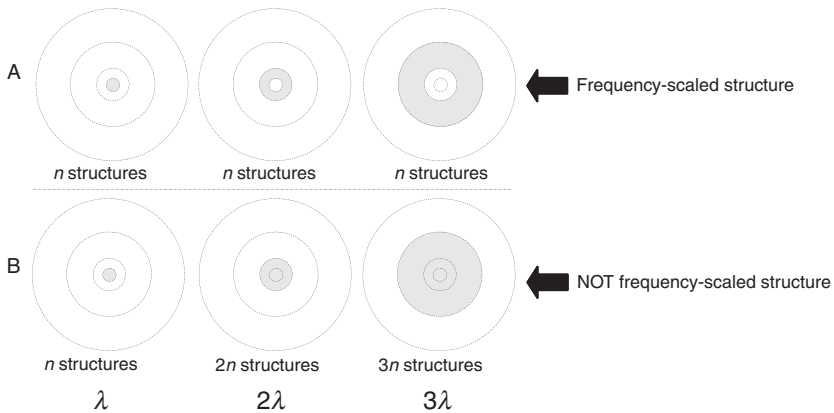


FIGURE 19-7 Different schemes of active regions: (a) Both the aperture size and structure are scaled with frequency. (b) Although aperture size is scaled with frequency, the structure is not.

wavelength, the structures within the active region are not. As frequency decreases, more and more structures are included in the active region and give rise to different antenna characteristics. On the other hand, the active region in the A scheme migrates outward and the structures in the active region are similar but scaled proportionally to the wavelength. Therefore, the A scheme is truly a frequency scaled design and should be adopted for UWB antenna designs. Note that the active region may or may not overlap the previous one as frequency varies. This depends on the design and the frequency step. Recall that the antenna is fed from the high frequency region, and there is strong mutual coupling between adjacent structures. To achieve scheme A, a special arrangement is necessary to reduce low-frequency radiation in high-frequency regions. Good examples can be found in the LPDA and spiral antenna designs shown in Figure 19-5. In the LPDA, the feeding polarity of the adjacent dipoles is reversed such that radiation from them is canceled at low frequencies where the phase delay from the length of these electrically short dipoles is negligible. In the two-arm spiral case, the adjacent arms in the high-frequency region are out of phase at low frequencies because the two arms are fed with 180° phase difference and the phase delay from the half-turn path difference is negligible at low frequencies. Another simpler way would be to simply bend the antenna arm back and forth as in the sinuous antenna case. More examples of Type-II UWB antennas can be found in Chapter 17.

19.6 THE DUAL-POLARIZATION UWB DIELECTRIC ROD PROBE (DRP) ANTENNA

Introduction

It is well known that a homogeneous dielectric rod can be used to guide electromagnetic waves. Unlike metallic waveguide, the lowest mode of a dielectric-rod waveguide (HE_{11} mode) does not have a cutoff frequency. As wavelength increases, more energy is

guided outside the rod. Most of the energy is guided inside the rod when the rod diameter becomes greater than the guided wavelength. This adaptive field distribution as a function of frequency makes a dielectric rod a good candidate for UWB antenna. First, radiation is generated from diffractions at the rod end. By controlling the materials and geometry at the end of the rod, different radiation, gain, pattern, and bandwidth can be achieved. A high-gain “polyrod” antenna was developed by using a long, tapered leaky-wave section at the end with a reported bandwidth of up to 2:1. A greater bandwidth can be achieved by adding multiple higher dielectric layers inside the rod to avoid pattern narrowing and higher waveguide modes. Recently, a double-layer, dual-polarization UWB dielectric rod probe antenna design was developed as a near-field probe with a bandwidth of more than 4:1. Several desirable features for a good near-field probe include²¹ (1) low radar cross section (RCS) area, desired to minimize the mutual coupling between a probe and the antenna under test (AUT); (2) dual-linear polarization with good polarization isolation; and (3) a broad beamwidth with a stable radiation phase center. These features are important to reduce probe correction error^{22,23} and to improve the accuracy of near-field to far-field transformation.

UWB DRP Configurations

A double-layer UWB dielectric rod probe antenna design that has a bandwidth of greater than 4:1 is illustrated in Figure 19-8. It consists of feeding, guiding, and radiating sections. A resistively terminated TEM launcher launches broadband electromagnetic waves into a guiding section where TEM waves are converted into HE_{11} mode. Finally, a specially designed radiation section produces stable and symmetric radiation patterns with a radiation center located near the tip. The final antenna is enclosed in a cylindrical conducting cylinder containing some absorbers to eliminate undesired radiations from feed and launchers. An absorber sleeve is placed around the exposed section of the rod with a 0.25-in air gap between rod and absorbers. Such an enclosure arrangement is to ensure that radiations only come from the radiation section.

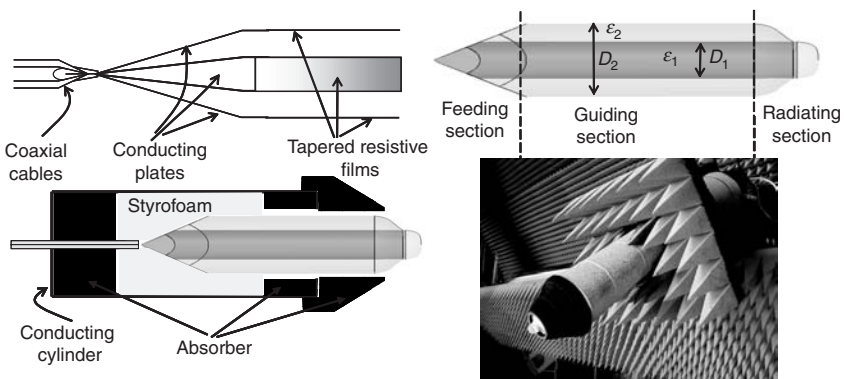


FIGURE 19-8 UWB dual-polarization dielectric rod probe antenna with more than 4:1 bandwidth and symmetric radiation patterns

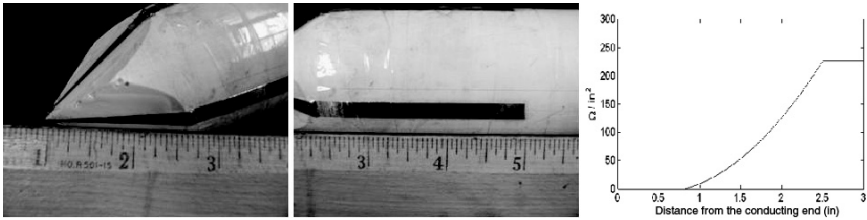


FIGURE 19-9 The feed section of a single-layer rod (2-in diameter) antenna with dual-polarization launchers (the scale on the ruler is inches)

Feed Section

The feed section shown in Figure 19-9 is responsible for launching UWB electromagnetic waves into the rod. This is achieved by a Type-II UWB launcher that consists of a pair of resistively terminated conducting plates that are mounted on the opposite surfaces of the feed section whose diameter is gradually increased to avoid excitation of higher mode. The geometry of the conducting plates is chosen to provide a desirable surge impedance of 100Ω , matched to the output impedance of the balun. The input impedance of the balun is 50Ω . The resistive tapering profile is plotted in Figure 19-9. The launcher design dictates the low-frequency limit of the rod antenna since little energy will be guided by the rod when the guided wavelength becomes much greater than the diameter of the rod. By adding a second launcher on the orthogonal plane, dual polarization can be easily achieved.

Guiding Section

The diameter and the dielectric constant (the outer layer, inner core) are chosen to be 2 inches and 1 inch, and $\epsilon_1 = 4$ and $\epsilon_2 = 6$, respectively. Figure 19-10 compares the calculated magnitude of electric fields guided by a single-layer (left) and double-layer (right) rod along the longitudinal cut at 2, 4, and 8 GHz. Both rods are 2-in diameter. The single-layer rod is lossless and has a dielectric constant of 4. The double-layer rod is also lossless and has dielectric constants of 4 and 6 in the outer layer and inner core (1-in diameter), respectively. In both cases, HE_{11} mode fields are excited from the bottom ends. For the single-layer rod, the amount of energy guided within the rod increases with frequency as expected. However, once most fields are guided within the rod, further frequency increase does not further reduce the size of the transverse field distribution. This can be seen at the top end of the single-layer rod at 8 GHz. If we let fields radiate from the top end, it will have a large effective radiation aperture (in terms of wavelength) at high frequencies, causing undesired pattern narrowing and lobes. As a comparison, the fields guided by the double-layer rod are mainly confined to the higher dielectric inner core at high frequencies. This would reduce the effective aperture size (in terms of wavelength) at high frequency. Therefore, it is conjectured that if thickness and dielectric constant of each layer of a multiple-layer rod are chosen properly, the frequency scaled radiation aperture should be obtainable and produce UWB gain and patterns.

Radiating Section

Figure 19-11 shows the geometry of the radiating section of the UWB rod antenna. Radiation arises from refraction and diffraction caused by the gradually reduced diameter.

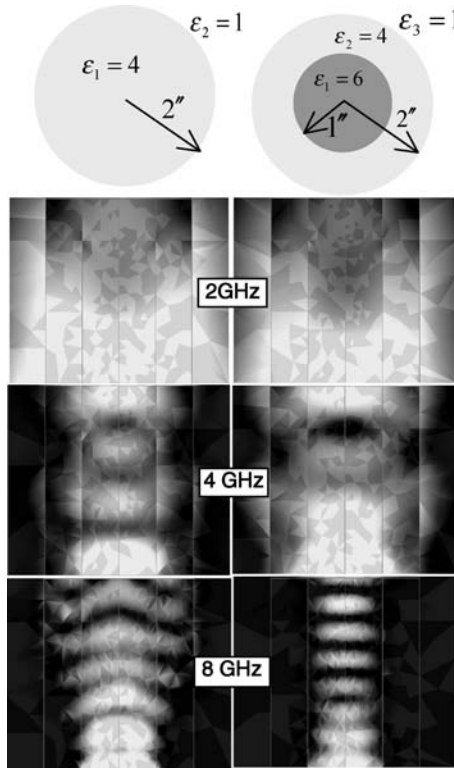


FIGURE 19-10 Comparison of field distribution (in the center longitudinal plane) of waves guided by single and double dielectric rods

We can also taper the dielectric constant down to produce radiations.²⁴ As the guided wave approaches the radiating section whose electrical size becomes much smaller than the guided wavelength, the fields can no longer be confined to the rod, causing radiation as illustrated in Figure 19-11. The length of the radiation section should be kept short

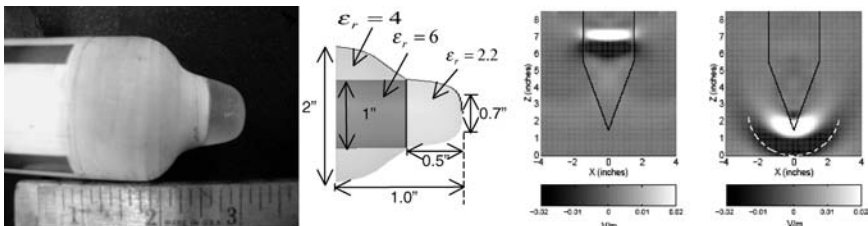


FIGURE 19-11 The radiating section of the UWB double-layer dielectric rod antenna

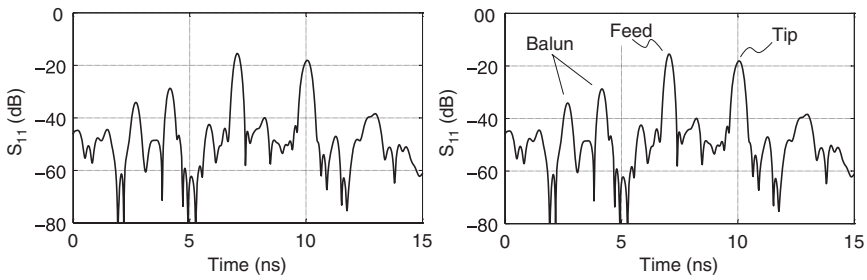


FIGURE 19-12 Measured VSWR (left) and return loss in time domain (right)

to avoid dispersion and pattern narrowing since shorter wavelengths can still be guided longer and radiate out near the top. However, a shorter length makes designing a smooth tapering more challenging. Internal reflections should also be minimized by a smooth tapering. Since the inner core has a higher dielectric constant, a small low-dielectric cover ($\epsilon_r = 2.2$) is added to provide a better transition into the air, thus reducing internal reflections at high frequencies.

Antenna Characteristics

Figure 19-12 plots the measured VSWR and time-domain return loss. The VSWR is < 2 within the 2~8-GHz band. The oscillatory behavior is caused by the interference between feed and tip reflections as clearly observed at the 7-ns and 10-ns position. The measured and simulated co- and cross-polarization gain in the boresight direction are plotted in Figure 19-13 (upper left), which shows a stable gain level of 2~5 dBi. The cross-polarization component appears to be approximately 20 dB below the copolarization component over the entire frequency range. The simulated and measured radiation patterns at 2, 4, and 8 GHz are also plotted in Figure 19-13. The designed goal of 60° beamwidth as a near-field probe is clearly met. The patterns also are relatively symmetrical in the E - and H -planes in the main beam region.

19.7 THE DUAL-POLARIZATION UWB DIELECTRIC HORN ANTENNA DESIGN

Introduction

A UWB dielectric horn antenna (DHA) was developed²⁵ for operating from 2 to 18 GHz. This is a Type-II UWB antenna because it utilizes absorbing materials to control undesired lateral waves and curved surface to reduce end diffraction such that there is only one source of radiation. The DHA's design objective is to provide dual-linear polarizations, wide and stable beamwidth, and symmetric radiation patterns. Figure 19-14 shows pictures of an assembled and partially assembled 2~18-GHz DHA antenna. The basic DHA configuration consists of a solid dielectric horn, a wave launcher, and a surrounding bulk absorber,

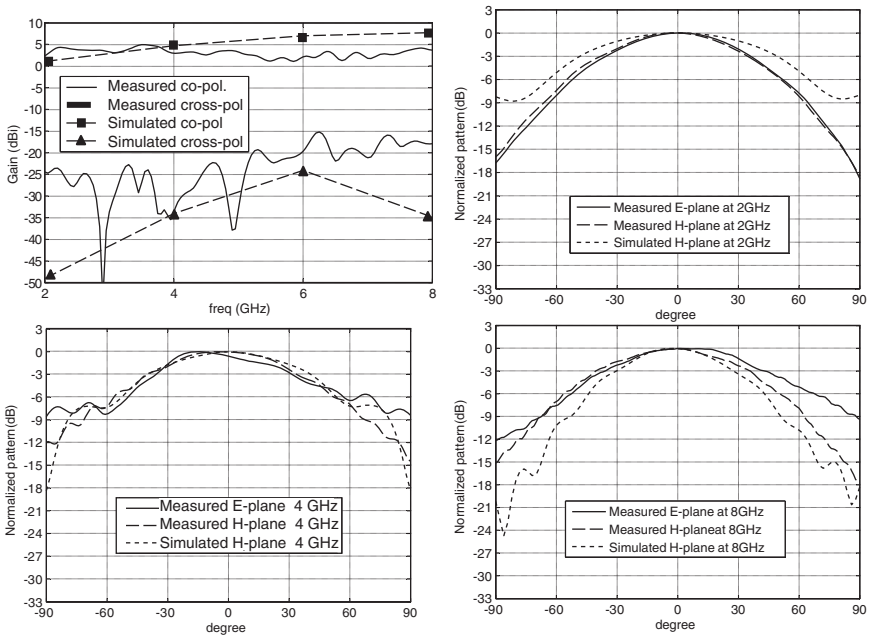


FIGURE 19-13 Measured and simulated realized gain and radiation patterns

as shown in Figure 19-14. Notice that this is not a dielectric loaded antenna. The dielectric horn is the primary radiator. This DHA design offers a much broader bandwidth (9:1) compared with most other dielectric horn antenna designs.²⁶

Wave Launchers and Feeding Arrangement

Similar to the previous dielectric rod antenna, each launcher arm begins with a triangular conducting plate and is followed by a tapered resistive strip. The surge impedance is a function of plate angle, horn angle, and the dielectric constant²⁷ and is chosen to be approximately

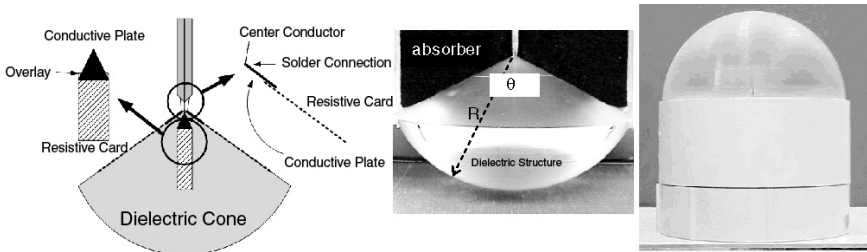


FIGURE 19-14 Basic configuration for the UWB DHA design

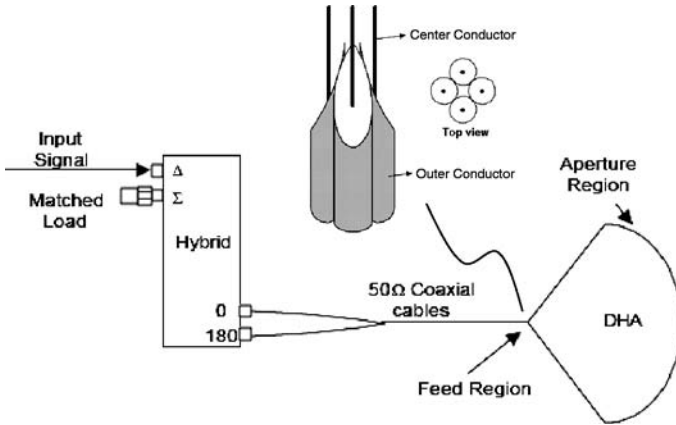


FIGURE 19-15 Balun and feeding cable arrangement of DHA design

100 Ω , matched to the output impedance of a balun that has 50- Ω input impedance (see Figure 19-15). The minimum length of the conducting plate is related to the antenna's low-frequency limit of antenna gain. The maximum length of the plate is related to the high-frequency limit of the radiation pattern.

The resistance profile of the resistively tapered strip is similar to the one shown in Figure 19-9 and is used to reduce launcher resonance. If the resistance value increases too rapidly, undesirable diffractions could occur along the tapering. On the other hand, if the resistance value increases too slowly, undesired reflections could occur at the end of the tapering due to short length. The width and length of the resistive film should be kept minimal to maintain dielectric-air boundary condition and to avoid the excitation of the cross-polarized fields. Of course, the utilization of such resistive loading is at the expense of lower radiation efficiency.

Surrounding Absorber

The surrounding absorber is used to suppress undesired lateral waves propagating along the sides. Such lateral waves exist when there is an interface between two media with different dielectric constants.²⁸ Figure 19-16 plots calculated field distribution of an

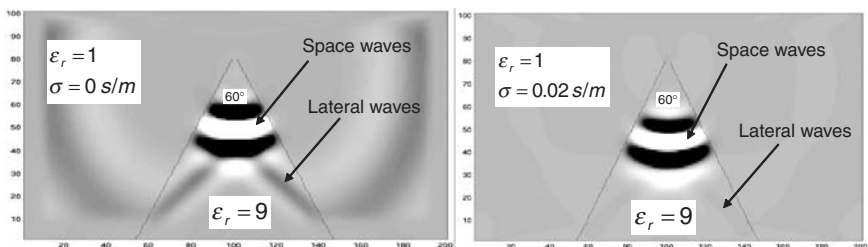


FIGURE 19-16 Propagation of impulse fields in DHA

impulse signal in a dielectric wedge with a dielectric constant of 9 and wedge angle of 60° . Without the external absorber (left figure), strong lateral waves are excited. Such lateral waves are easily identified by the angle, $\sin^{-1}(\epsilon_{\text{external}}/\epsilon_{\text{internal}})$, their wavefronts made with the dielectric interface. The presence of these lateral waves eventually will also interfere with radiation from the space wave, causing undesired variation in the gain and pattern with frequency and angle.²⁹ Since lateral waves extend to the external medium, they can be suppressed by applying an external absorber. It is evident that lateral waves have been largely suppressed with only little impact on space waves. Also notice the much weaker external fields, which are beneficial in suppressing side and back lobes.

Dielectric Horn

The end of the dielectric horn has a spherical surface with its spherical center being at the feedpoint. A low-loss dielectric material, polymethyl-methacrylate (acrylic), is chosen for its low cost and easy fabrication. Acrylic also has good ultraviolet (UV) and thermal stability, low water absorption, and high dielectric strength. Its typical refractive index is approximately 1.48 over the frequency range of interest (2–18 GHz). Although choosing a material with a higher dielectric constant could reduce external waves and reduce antenna dimensions, it also increases undesired internal reflection at the spherical aperture and diffractions at edges. Such reflections could be reduced via the utilization of proper dielectric tapering along the radial direction for a smooth impedance transition.

The horn angle should be chosen based on desired beamwidths. Our empirical data showed the -10 -dB beamwidth was approximately $10^\circ\sim 13^\circ$ less than the horn angle (less than 70°). A larger horn angle produces a wider beamwidth but requires wider launcher arms to maintain the same input impedance. This may be problematic since it can increase the cross coupling in a dual-polarization design. A large horn angle is also more susceptible to higher-order modes that could be excited by slight asymmetry in the two launcher arms. Figure 19-17 plots snapshots of computed pulsed fields in a DHA with a horn angle of 130° and dielectric constant of 2.2. This simulation adopts a realistic absorber with a dielectric constant of 1.8 and conductivity of 0.22 S/m at 10 GHz . It is observed that the fields propagate outward spherically without experiencing magnitude and phase distortion. The fields are also confined within the horn. To further reduce undesired diffractions at the edge of the spherical end, an improved design adopts smoothly curved sides, as illustrated in Figure 19-18. Notice the pyramidal geometry near the feed region for achieving improved isolation between

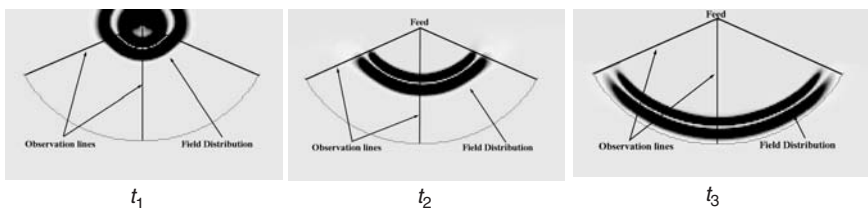


FIGURE 19-17 Propagation of pulsed fields in a dielectric horn

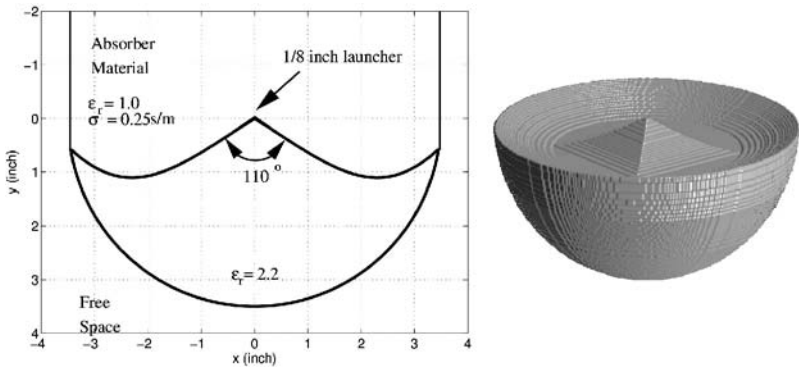


FIGURE 19-18 Dielectric horn geometry of improved DHA design for reducing edge diffraction

orthogonal polarizations in a dual-polarization design compared with the circular cone geometry (see Figure 19-14).

Antenna Characteristics

In this section, measured data for a DHA designed for 2–18-GHz dual-polarization operations with a beamwidth of approximately 60° are presented. This DHA is pictured in Figure 19-14. The radius and angle of the horn are 3.5 in and 110°, respectively. Figure 19-19 shows the measured VSWR, which is less than 1.5 for most of the band and below 2.0 over

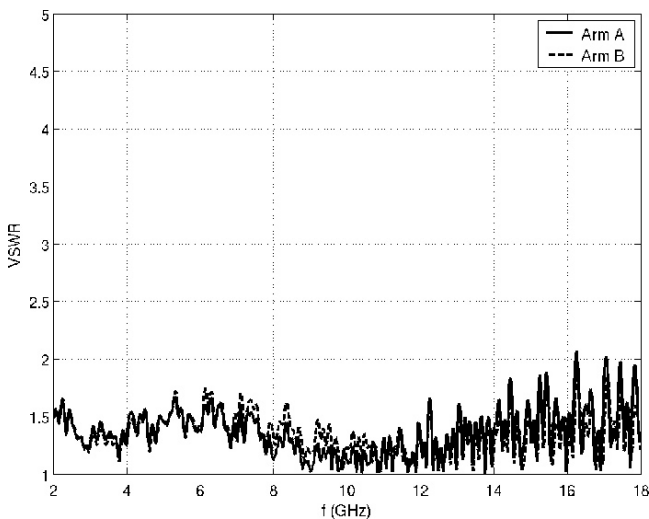


FIGURE 19-19 Measured VSWR from 2 to 18 GHz

the entire band. Although this is somewhat expected due to the utilization of resistively loaded launchers, this result does indicate low reflection from spherical aperture, the main reason for not using high-dielectric material. The antenna efficiency of this DHA is around 10 percent and 27 percent at 2 GHz and 18 GHz, respectively. The solid and dashed lines are used to distinguish the two orthogonal polarizations.

Figure 19-20 plots measured boresight gain (co- and cross-polarization components) for two orthogonal polarizations. The gain level increases approximately 5 dB from 2 to 18 GHz. The small rapid fluctuations are clutter from the measurement facility. The low gain level is due to the utilization of the absorber and the tapered resistive strips. Unfortunately, the absorber is necessary in the current design to eliminate the lateral waves. However, a high gain antenna is not a critical need in many applications as in the indoor ranges. To adapt the antenna to other applications where a higher gain level is desired, other launcher designs that could avoid launching lateral waves are possible and are being investigated. The copolarization-to-cross-polarization isolation decreases monotonically from 25 dB at 2 GHz to 15 dB at 18 GHz. This behavior of decreasing isolation is caused by slight misalignments in placing the launcher arms or feeding wires.

Figure 19-21 plots the normalized measured far-field patterns, which also agree well with simulated data²⁵ (not shown here). These patterns show good *E*- and *H*-plane symmetry as well as stable beamwidth. It is observed that the beamwidth below 4 GHz becomes significantly wider since the cross sectional size at the effective end of the launcher is small compared with wavelength such that HE_{11} mode is only loosely guided. Thus, for our application, the antenna has a 9:1 bandwidth. Only the radiation patterns for one of the two orthogonal polarizations are shown here since both polarizations have similar patterns.

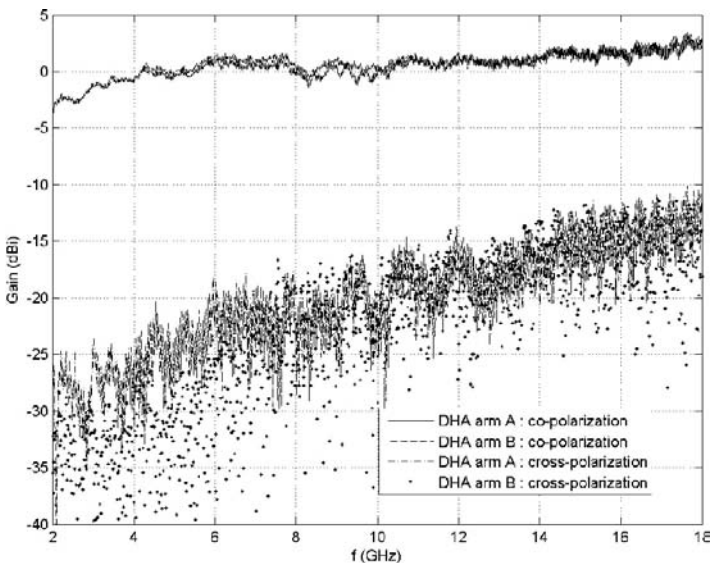


FIGURE 19-20 Measured gain from 2 to 18 GHz

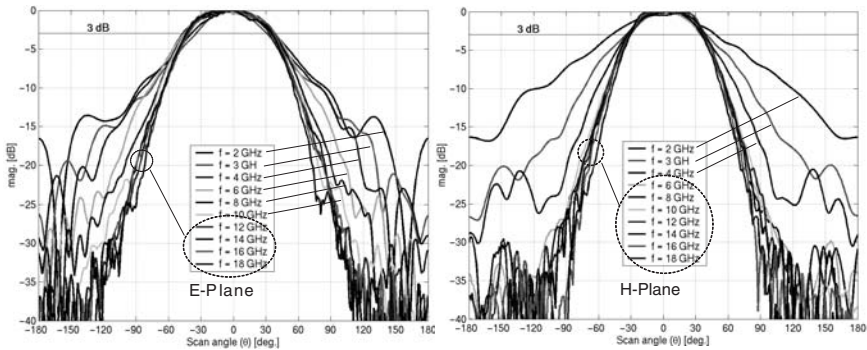


FIGURE 19-21 Normalized measured far-field radiation patterns in the *E*-plane (left) and *H*-plane

REFERENCES

1. N. P. Agrawal, G. Kumar, and K. P. Ray, "Wide-Band Planar Monopole Antennas," *IEEE Trans. Antennas Propagat.*, vol. 46, no. 2 (1998): 294–295.
2. Z. N. Chen, M. Y. W. Chia, and M. J. Ammann, "Optimization and Comparison of Broadband Monopoles," *IEE Proc. Microwave Antennas Propagat.*, vol. 150, no. 6 (2003): 429–435.
3. K. G. Thomas, N. Lenin, and R. Sivaramakrishnan, "Ultrawideband Planar Disc Monopole," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 4 (April 2006): 1339–1341.
4. R. G. Kouyoumjian and P. H. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *Proc. IEEE*, vol. 62 (November 1974): 1448–1461.
5. T. T. Wu and R. W. P. King, "The Cylindrical Antenna with Nonreflecting Resistive Loading," *IEEE Trans. Antennas Propagat.*, vol. 13 (May 1965): 369–373.
6. L. C. Shen and R. W. P. King, "Corrections to 'The Cylindrical Antenna with Nonreflecting Resistive Loading,'" *IEEE Trans. Antennas Propagat.*, vol. 13 (November 1965): 998.
7. M. Kanda, "A Relatively Short Cylindrical Broadband Antenna with Tapered Resistive Loading for Picoseconds Measurements," *IEEE Trans. Antennas Propagat.*, vol. AP-26 (May 1978): 439–447.
8. K. P. Esselle and S. S. Stuchly, "Pulse-Receiving Characteristics of Resistively Loaded Dipole Antennas," *IEEE Trans. Antennas Propagat.*, vol. 38, no. 10 (October 1990): 1677–1683.
9. L.-C. T. Chang and W. D. Burnside, "An Ultrawide-Bandwidth Tapered Resistive TEM Horn Antenna," *IEEE Trans. Antennas Propagat.*, vol. 48 (December 2000): 1848–1857.
10. See Chapter 10.
11. R. Janaswamy and D. H. Schaubert, "Analysis of the Tapered Slot Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-35 (September 1987): 1058–1064.
12. E. Gazit, "Improved Design of the Vivaldi Antenna," *IEE Proc. Microwave Antennas Propagat.*, vol. 135, no. 2 (1988): 89–92.
13. D. Heedy and W. D. Burnside, "An Aperture-Matched Compact Range Feed Horn Design," *IEEE Trans. Antennas Propagat.*, vol. 33, no. 11 (November 1985): 1249–1255.
14. K. Y. A. Lai, A. L. Sinopoli, and W. D. Burnside, "A Novel Antenna for Ultrawide-Band Applications," *IEEE Trans. Antennas Propagat.*, vol. 40 (July 1992): 755–760.
15. K. L. Walton and V. C. Sundberg, "Broadband Ridged Horn Design," *Microwave J.* (March 1964): 76–101.
16. C. Bruns, P. Leuchtmann, and R. Vahldieck, "Analysis and Simulation of a 1–18-GHz Broadband Double-Ridged Horn Antenna," *IEEE Trans. on Electromagnetic Compatibility*, vol. 45 (February 2003): 55–60.

17. J. L. Kerr, "Short Axial Length Broad-Band Horns," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (September 1973): 710–714.
18. J. D. Dyson, "The Equiangular Spiral Antenna," *IRE Trans. Antennas Propagat.*, vol. 7 (April 1959): 181–187.
19. B. R.-S. Cheo, V. H. Rumsey, and W. J. Welch, "A Solution to the Frequency-Independent Antenna Problem," *IRE Trans. Antennas Propagat.* (November 1961): 527–534.
20. D. Isbell, "Log Periodic Dipole Arrays," *IEEE Trans. Antennas Propagat.*, vol. 8, no. 3 (May 1960): 260–267.
21. J. H. Richmond and T. E. Tice, "Probes for Microwave Near-Field Measurements," *IRE Trans. Microwave Theory Tech.*, vol. MTT-3 (April 1955): 32–34.
22. D. T. Paris, W. M. Leach, Jr., and E. B. Joy, "Basic Theory of Probe-Compensated Near-Field Measurements," *IEEE Trans. Antennas Propagat.*, vol. 26 (May 1978): 373–378.
23. E. B. Joy, W. M. Leach, Jr., G. P. Rodrigue, and D. T. Paris, "Applications of Probe-Compensated Near-Field Measurements," *IEEE Trans. Antennas Propagat.*, vol. 26 (May 1978): 379–389.
24. C.-C. Chen, K. R. Rao, and R. Lee, "A Tapered-Permittivity Rod Antenna for Ground Penetrating Radar Applications," *J. Appl. Geophys.*, vol. 47 (3-4) (September 2001): 309–316.
25. K. W. Lee, C.-C. Chen, and R. Lee, "UWB Dual-Linear Polarization Dielectric Horn Antennas as Reflector Feeds," *Trans. Antennas Propagat.*, vol. 55, no.3 (March 2007): 798–804.
26. C. Salema, C. Fernandes, and R. K. Jha, *Solid Dielectric Horn Antenna* (Norwood, MA: Artech House, Inc., 1998).
27. N. Venkatarayalu, C.-C. Chen, F. L. Teixeira, and R. Lee, "Numerical Modeling of Ultra-Wideband Dielectric Horn Antennas Using FDTD," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 5 (May 2004): 1318–1323.
28. J. R. Wait, *Electromagnetic Waves in Stratified Media* (New York: IEEE Press, 1996).
29. S. J. Radzevicius and J. J. Daniels, "GPR H-Plane Antenna Patterns for a Horizontal Dipole on a Half Space Interface," GPR Proceedings, Gold Coast, Australia, 2000.

Chapter 20

Phased Arrays

R. C. Hansen

R. C. Hansen, Inc.

CONTENTS

20.1	INTRODUCTION.....	20-2
20.2	THEORY OF PHASED ARRAYS	20-3
20.3	PATTERN CONTROL	20-15
20.4	RADIATING ELEMENT DESIGN AND MUTUAL COUPLING	20-21
20.5	FEED DESIGN FOR PHASED ARRAYS	20-26

20.1 INTRODUCTION

In this chapter, the theoretical and practical consideration involved in the design of phased array antennas is given. The intent is to provide the antenna designer with an overview of the basic design concepts of phased arrays and the means to determine the proper antenna parameters for a given application. This chapter is an updating of that prepared for the previous edition by Ray Tang and Richard Burns. They were associated with Hughes Aircraft Co., Fullerton, California.

Basic Concept of Phased Arrays

Basically, the phased array antenna is composed of groups of individual radiators (elements), which are distributed and oriented in a linear or two-dimensional spatial configuration. The amplitude and phase excitations of each element can be individually controlled to form a radiated beam of any desired shape (from a directive pencil beam to a fan beam shape) in space. The position of the beam in space is controlled electronically by adjusting the phase of the excitation signals at the individual elements. Hence, beam scanning is accomplished with the antenna aperture remaining fixed in space without the involvement of mechanical motion in the scanning process. In other words, this is scanning without inertia.

Advantages of Phased Arrays

The capability of rapid and accurate beam scanning in microseconds permits the system to perform multiple functions, either interlaced in time or simultaneously. An electronically-steered array radar is able to track a large number of targets and illuminate some of these targets with radio frequency (RF) energy, for the purpose of guiding missiles toward them. This form of radar can perform complete hemispherical searches, with automatic target selection and hand-over to tracking. It may even act as a communication system, directing high gain beams toward distant receivers and transmitters. Complete flexibility is possible. Search and track rates may be adjusted to best meet the particular situation within the limitations set by the total use of time. The antenna beamwidth may be changed electronically, by means of phase spoiling, to cover certain areas more rapidly but with less gain. Very high powers may be generated from a multiplicity of generators distributed across the aperture. Electronically controlled phased-array antennas can give the flexibility needed to perform all the various functions in a way best suited to the specific task at hand. The functions can be programmed rapidly and accurately with digital beam steering computers.

Overview of the Chapter

The first part of this chapter describes the theory of phased-array antennas, with emphasis on the basic relationships between the array performance, the excitation coefficients (amplitude and phase) of each element, and the physical parameters of the antenna (see Sections 20.2 and 20.3). These relationships are the design formulas, which an antenna designer can use to determine the required array parameters that satisfy a given set of system requirements. After determining the array parameters such as the number of required elements, the spacing between the elements, and the size of the radiating aperture, the next step in the design would be the selection of the type of element, phaser, and beamforming

feed network to implement the required excitations at the radiating aperture physically. A detailed discussion of the various types of elements and feed networks that are commonly used is given in Sections 20.4 and 20.5, respectively. Corresponding to the various types of feed networks, the frequency-bandwidth characteristics of the phased-array antenna are described in Section 20.5.

20.2 THEORY OF PHASED ARRAYS

We shall begin the discussion of array theory with the simplest configuration possible: an array of isotropic elements equally spaced along a line (linear in a geometric sense). It will be seen that the more complex array theory can be framed as an extension of this simple case. For example, two-dimensional arrays are often constructed of interconnected linear arrays. From the linear array discussion, we will then proceed to the theory of two-dimensional arrays. The theory of arrays of isotropic radiators will then be extended and modified to apply to arrays of real elements.

To simplify and condense the material to be presented, the following assumptions will be adhered to in this discussion:

1. For practical purposes the far-field region of the antenna, commonly defined by ranges R , satisfying

$$R \geq \frac{2L^2}{\lambda} \quad (20-1)$$

where L is the largest dimension of the antenna and λ is the operating wavelength of interest. In this region, the antenna pattern is insensitive to range except for a scale factor of $1/R^2$ in power. This scale factor is usually ignored.

2. The reciprocity theorem¹ will be used extensively to justify analysis of array far-fields from either the transmitting or the receiving viewpoint, as convenience dictates and usually without explicit mention. By virtue of this theorem, the pattern will be the same in either case if no nonreciprocal devices are used. When such devices are used, one can usually perform pattern analyses which neglect their presence.
3. For simplicity, interest will be restricted to the responses of arrays to continuous wave (CW) signals.

Radiation Pattern of Linear Arrays

The theory of linear arrays is described in detail by Silver¹ and Schelkunoff.² Let us consider the elementary array of Figure 20-1, consisting of N isotropic elements, equally spaced at a distance d apart. On receive, if a plane wave is incident upon the array from a direction making an angle θ with the array normal, the current in the n th element will be of the form

$$i'_n = A e^{jnkd \sin \theta} \quad (20-2)$$

where A is a complex constant related to the instantaneous amplitude and phase of the plane wave and k is the wave number:

$$k = \frac{2\pi f}{c} = \frac{2\pi}{\lambda} \quad (20-3)$$

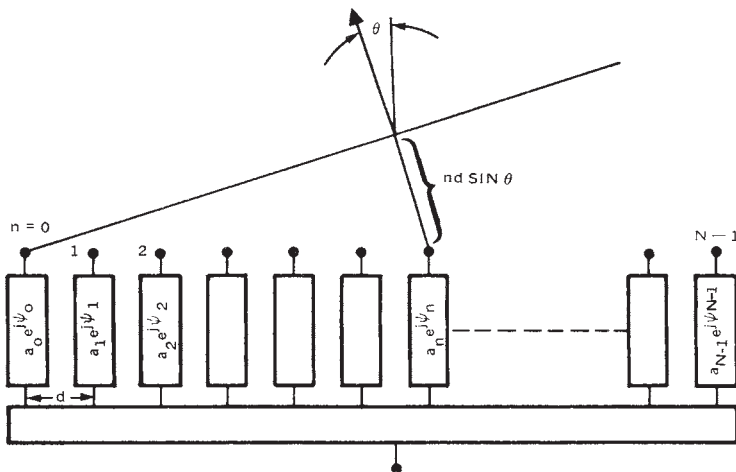


FIGURE 20-1 Basic linear array configuration

where

f = operating frequency
 c = velocity of light

Eq. 20-1 shows that the current in the n th element leads that in the $(n + 1)$ th element by a phase shift given by $\Delta\psi = kd \sin \theta$. This phase shift corresponds to the difference in time of arrival τ of the plane wavefront of $\tau = d/c \sin \theta$. If we place a control element behind each radiator, as indicated in Figure 20-2, with a transfer coefficient for the n th element given by

$$\frac{i''_n}{i'_n} = a_n e^{j\psi_n} \quad (20-4)$$

where a_n and ψ_n are the *real* current gain and phase shift of the control element, respectively, the summing network produces an output

$$E_a(q) = \sum_{n=0}^{N-1} a_n e^{j(\psi_n + nkd \sin \theta)} \quad (20-5)$$

where we now neglect the constant A of Eq. 20-2. This relationship gives the response of the array of Figure 20-2 to a signal arriving from a direction θ in terms of the set a_n 's and ψ_n 's. The set of coefficient a_n is usually called the *array amplitude taper*, while the ψ_n 's are called the *phase taper*. The expression of Eq. 20-5 is called the *array factor*. To combine the received signals from all the radiators in phase to produce a maximum response in the scan direction of θ_0 , the ψ_n 's must have the form

$$\psi_n = -nkd \sin \theta_0 \quad (20-6)$$

This expression shows that the required phase taper across the array aperture is a linear taper (constant phase differential between adjacent radiators). On transmit, when the phases

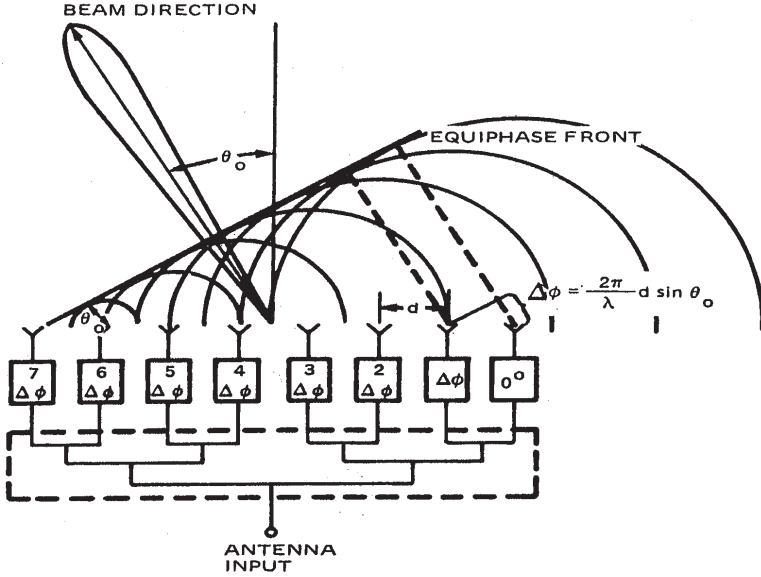


FIGURE 20-2 Beam-steering concept using phasers (phase shifters) at each radiating element

of the control elements such as phasers are set to the phase taper of Eq. 20-6, the signals radiated from all the elements will add up in phase to produce a main beam in the direction of θ_0 . Hence, the array factor of Eq. 20-5 is the same for both transmit and receive. Substituting Eq. 20-6 into Eq. 20-5, we have for the array factor

$$E_a(\theta) = \sum_{n=0}^{N-1} a_n e^{jnk d (\sin \theta - \sin \theta_0)} \quad (20-7)$$

For the special case of a uniformly illuminated array, $a_n = 1$ for all n , the array factor for an N -element array becomes

$$E_a(\theta) = \frac{\sin \left[N\pi \frac{d}{\lambda} (\sin \theta - \sin \theta_0) \right]}{N \sin \left[\pi \frac{d}{\lambda} (\sin \theta - \sin \theta_0) \right]} \quad (20-8)$$

Grating Lobes Versus Element Spacing

The array factor of Eq. 20-7 can also be expressed in terms of the variable $v = \sin \theta$ as follows:

$$E_a(v) = \sum_{n=0}^{N-1} a_n e^{jnk d (v - v_0)} \quad (20-9)$$

where the beam direction v_0 is related to the differential phase $\Delta\psi$ by $\Delta\psi = -kdv_0$. It is obvious that $E_a(v_0)$ and $E_a(\theta)$ are related by a one-to-one mapping in the region $|v| \leq 1$, which is often referred to as the *visible space* corresponding to the real angles of θ . It is also apparent that $E_a(v)$ is a periodic function of v of period

$$\frac{2\pi}{kd} = \frac{1}{d/\lambda} = \frac{\lambda}{d} \quad (20-10)$$

and that Eq. 20-9 is in the form of a Fourier series representation, which is readily analyzable and easy to visualize. The maxima of $E_a(v)$ occur whenever the argument of Eq. 20-9 is a multiple of 2π , for example, $kd(v - v_0) = 2i\pi$, where $i = 0, \pm 1, \pm 2, \dots$, or

$$v_i - v_0 = \frac{i}{d/\lambda} \quad (20-11)$$

When $v_i = v_0$ or $i = 0$, this maximum is generally referred to as the *principal lobe* (PL) or main beam, and the other maxima are known as the *grating lobes* (GL) from the corresponding phenomena with optical gratings. In the design of phased arrays, it is imperative that the grating lobes be eliminated within the visible space since these lobes reduce the power in the main beam and thus reduce the antenna gain, and introduce unwanted signals and noise. This means that the element spacing d must be chosen to avoid the grating lobes over the range of v from -1 to $+1$. When the main beam is scanned to v_0 , the closest grating lobe to the visible space is located at $v_i = v_0 - \lambda/d$ (see Figure 20-3). This grating lobe will just appear in visible space (at the endfire direction of the array) when $v_0 - \lambda/d = -1$, or

$$\frac{d}{\lambda} = \frac{1}{1 + \sin |\theta_0|} \quad (20-12)$$

Thus, the element spacing criterion stated in terms of the desired maximum scan angle $\theta_{0\max}$ is

$$\frac{d}{\lambda} < \frac{1}{1 + \sin |\theta_{0\max}|} \quad (20-13)$$

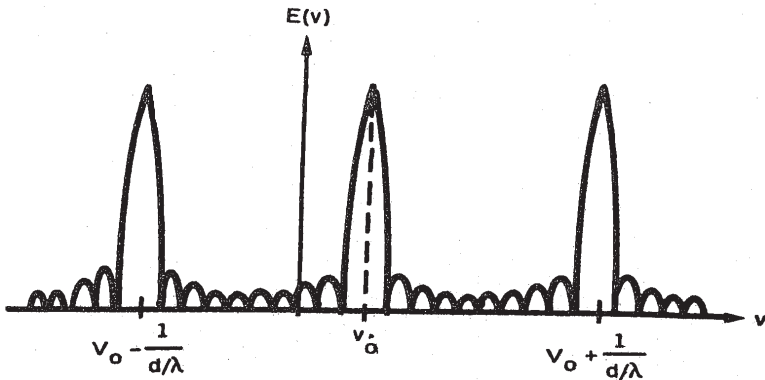


FIGURE 20-3 Grating-lobe formation

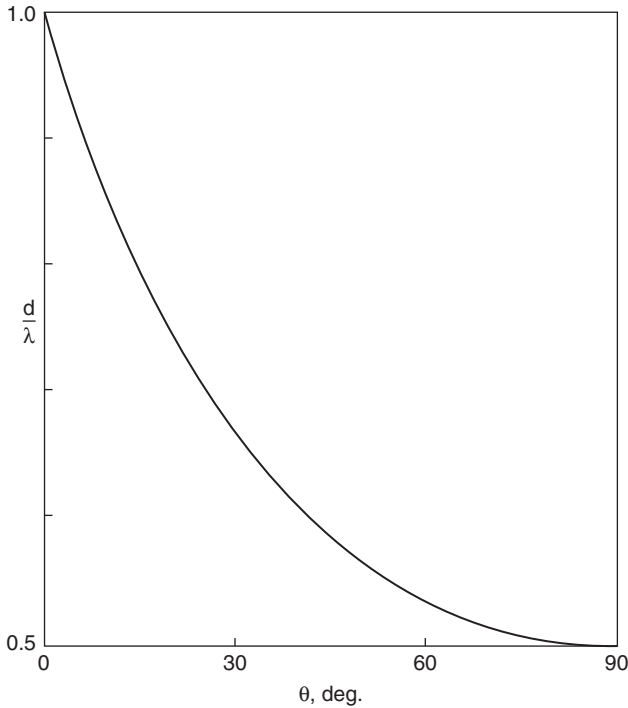


FIGURE 20-4 Scan angle versus element spacing for GL appearance

Figure 20-4 is a plot of Eq. 20-13. Note that GL appear when element spacing is sufficient; scan only moves the GL. For small arrays, the grating lobe null may be positioned at 90° :

$$\frac{d}{\lambda} = \frac{N - \sqrt{1 + \beta^2}}{N(1 + \sin \theta)} \quad (20-14)$$

where $\beta = 0$ for uniform excitation, or B for Taylor One-Parameter distribution. The analysis of Eq. 20-14 is due to David Munger and Richard Phelan.

Sub-Arrays and Quantization Lobes

Linear or planar arrays for limited scan may employ sub-arrays, where a small group of elements (sub-arrays) are fed together. Thus expensive hardware such as phasers, preamps, power amps, and combiners is minimized. When scanned, an array of contiguous sub-arrays has a stair-step phase, with one step over each sub-array. This gives rise to *quantization lobes* (QL). These are located at approximately the grating lobe angles, except that the sub-array width W is used in Eq. 20-11. QL amplitudes are given approximately by $\text{sinc}(\pi W/\lambda \sin \theta)$. For example, for $W/\lambda \sin \theta = .3$ the QL amplitude is -12.6 dB, while for $W/\lambda \sin \theta = .5$ the QL is -3.9 dB. Note that unlike GL, which appear whenever element spacing is adequate, QL appear only when the beam is scanned away from broadside.³

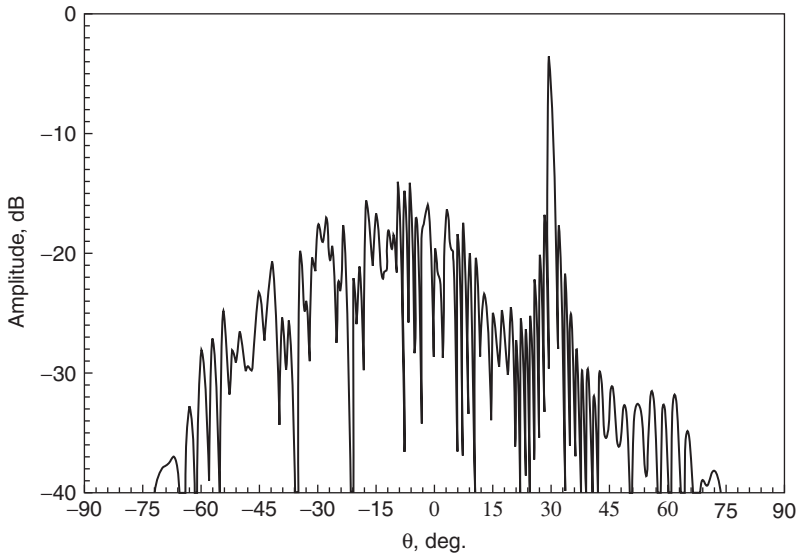


FIGURE 20-5 100 element array, 2 element subarray, and 30° scan

Quantization lobes can be significantly attenuated by clever use of random numbers, as shown by Smith and Guo.⁴ Random numbers are added to the drive of the digital phasers; then the same set, truncated to the number of phaser bits used, is subtracted. If the random numbers are chosen properly, the QL are depressed. Figure 20-5 shows an array pattern with two element sub-arrays at 30° scan. When the pseudo-randomization technique is used, the QL are reduced.⁵

Radiation Pattern of Planar Arrays

To provide beam scanning in two angular dimensions, a planar array of radiating elements must be used. In a spherical coordinate system, the beam position is defined by the two coordinates θ and ϕ , as shown in Figure 20-6. Also shown in Figure 20-6 is the layout of the element lattice in the planar array. For a rectangular lattice, the mn th element is located at $x_m = md_x$ and $y_n = nd_y$. For a triangular lattice, the element grid can be thought of as a rectangular grid in which every other element has been omitted. In this case, only every other value of mn contains an element. The element locations can be defined by requiring that the $(m + n)$ be even.

The array factor of a two-dimensional array may be calculated by summing the vector contribution of each element in the array at each point in space. The array factor can be written in terms of the directional cosines, $\cos \alpha_x$ and $\cos \alpha_y$, of the spherical coordinate system as follows:

$$E_a(\cos \alpha_x, \cos \alpha_y) = \sum_m \sum_n i_{mn} e^{jk(md_x \cos \alpha_x + nd_y \cos \alpha_y)} \quad (20-15)$$

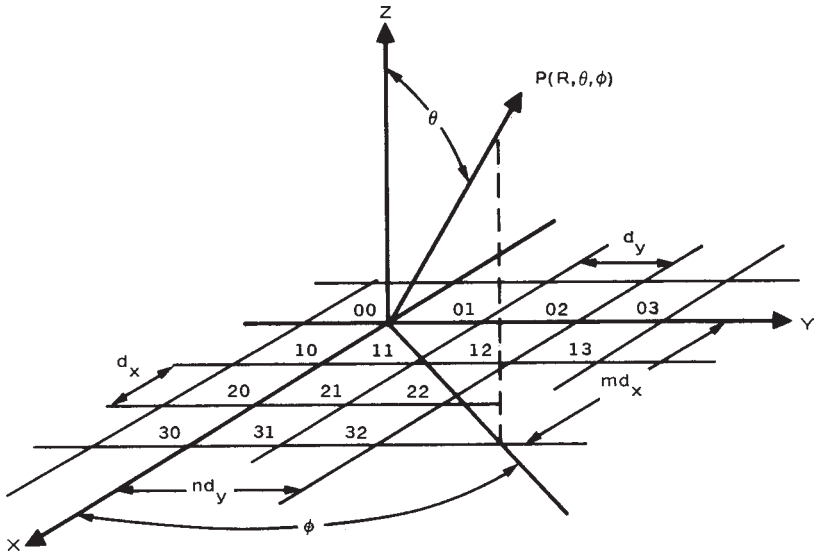


FIGURE 20-6 Element geometry of a two-dimensional array

where

$$\begin{aligned}\cos \alpha_x &= \sin \theta \cos \phi \\ \cos \alpha_y &= \sin \theta \sin \phi\end{aligned}\quad (20-16)$$

For a uniformly illuminated ($i_{mn} = 1$) rectangular array, we have

$$E_a(\cos \alpha_x, \cos \alpha_y) = \sum_{m=-(M-1)/2}^{(M-1)/2} e^{jkmd_x \cos \alpha_x} \sum_{n=-(N-1)/2}^{(N-1)/2} e^{jknd_y \cos \alpha_y} \quad (20-17)$$

Each sum can be evaluated, producing a result analogous to Eq. 20-8 for a uniformly illuminated linear array:

$$E_a(\cos \alpha_x, \cos \alpha_y) = \left[\frac{\sin \left(\pi M \frac{d_x}{\lambda} \cos \alpha_x \right)}{M \sin \left(\pi \frac{d_x}{\lambda} \cos \alpha_x \right)} \right] \left[\frac{\sin \left(\pi N \frac{d_y}{\lambda} \cos \alpha_y \right)}{N \sin \left(\pi \frac{d_y}{\lambda} \cos \alpha_y \right)} \right] \quad (20-18)$$

Beam scanning with planar arrays is accomplished by linear phasing along both array coordinates. To scan the beam to the angular position corresponding to the directional cosines $\cos \alpha_{x0}$ and $\cos \alpha_{y0}$, a linear phase taper is introduced at each element so that the excitation at the mn th element is given by

$$i_{mn} = a_{mn} e^{j(kmd_x \cos \alpha_{x0} + knd_y \cos \alpha_{y0})} \quad (20-19)$$

where

$kd_x \cos \alpha_{x0}$ = element-to-element phase shift in the x direction

$kd_y \cos \alpha_{y0}$ = element-to-element phase shift in the y direction

This form of steering phase indicates that the phase of the m th element is the sum of a row phase $mkd_x \cos \alpha_{x0}$ and a column phase $nkdy \cos \alpha_{y0}$.

The array factor of a rectangular planar array of M by N elements is then given by

$$E_a(\cos \alpha_x, \cos \alpha_y) = \sum_m \sum_n a_{mn} e^{jk[md_x(\cos \alpha_x - \cos \alpha_{x0}) + nd_y(\cos \alpha_y - \cos \alpha_{y0})]} \quad (20-20)$$

Element Spacing and Lattice of Planar Arrays

As in the case of the linear array, the array factor of the planar array Eq. 20-18 has an infinite number of grating lobes in the directional cosine space. For example, the maxima of E_a occurs whenever the argument of Eq. 20-18 is a multiple of 2π . Since there is a one-to-one correspondence between the directional cosine space ($\cos \alpha_x$ and $\cos \alpha_y$ space) and the visible space (θ and ϕ space within the boundary defined by $\cos^2 \alpha_x + \cos^2 \alpha_y = 1$), the number of grating lobes that can be projected from the directional cosine space into the visible space depends upon the parameters d_x/λ and d_y/λ . To avoid the formation of grating lobes in the visible space, the element spacings d_x/λ and d_y/λ must be chosen so that there is only one maximum from Eq. 20-18, namely, the main beam, in the visible space (real space). In the planar array, the element lattice and spacing can be chosen to shape the grating lobe contour (location pattern of grating lobes) to fit the required scanning volume, so that the total required number of elements in the planar array is minimized. To accomplish this optimization, it is more convenient to plot the position of the grating lobes when the main beam is phased for broadside and observe the motion of these lobes as the beam is scanned. Figure 20-7 shows the grating lobe locations for both rectangular and triangular spacings. For a rectangular lattice, the grating lobes are located at

$$\begin{aligned} \cos \alpha_x - \cos \alpha_{x0} &= \pm \frac{\lambda}{d_x} \cdot p \\ p, q &= 0, 1, 2, \dots \\ \cos \alpha_y - \cos \alpha_{y0} &= \pm \frac{\lambda}{d_y} \cdot q \end{aligned} \quad (20-21)$$

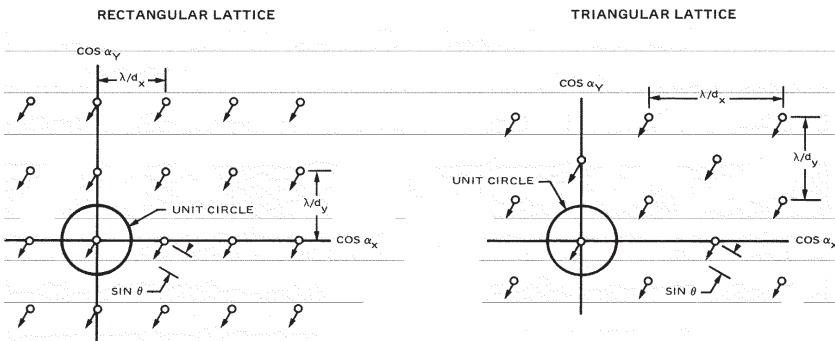


FIGURE 20-7 Grating-lobe locations for rectangular and triangular element lattices

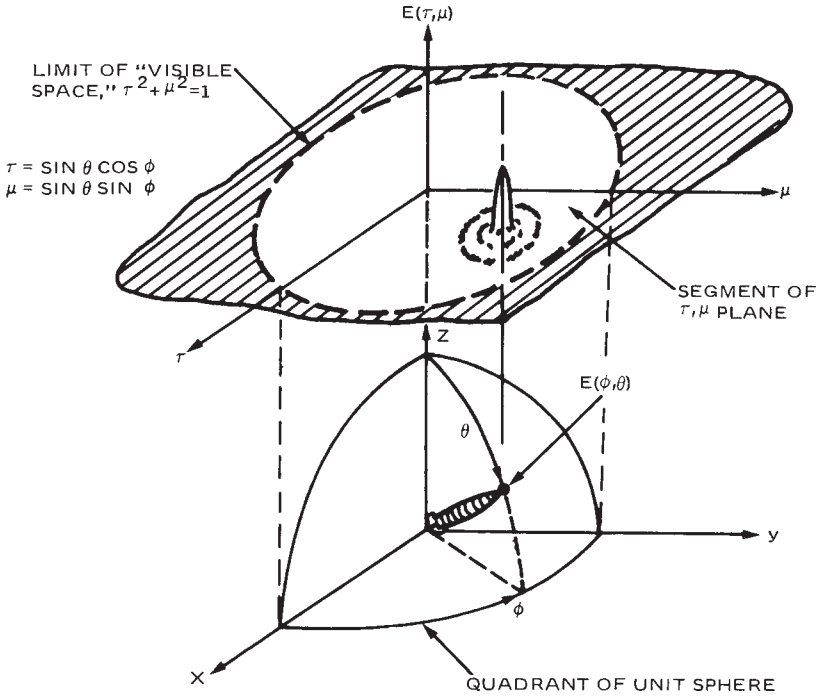


FIGURE 20-8 Construction of array factor $E(\phi, \theta)$ by projection of array function $E(\tau, \mu)$ onto unit sphere

The grating lobe pattern of Figure 20-7 must be mapped onto the surface of the unit sphere, as shown in Figure 20-8, to give a true spatial distribution; therefore, only the portion of the pattern of Figure 20-7 inside a unit circle centered at $\cos \alpha_x = \cos \alpha_y = 0$ lies in visible space. The lobe at $p = q = 0$ is the main beam. For a conical scan volume, the triangular grid is more efficient for the suppression of grating lobes than a rectangular grid,⁶ so that for a given aperture size fewer elements are required. If the triangular lattice contains elements at md_x and nd_y , where $m + n$ is even, then the grating lobes are located at

$$\cos \alpha_x - \cos \alpha_{x0} = \pm \frac{\lambda}{2d_x} \cdot p$$

$$p, q = 0, 1, 2, \dots \quad (20-22)$$

$$\cos \alpha_y - \cos \alpha_{y0} = \pm \frac{\lambda}{2d_y} \cdot q$$

where $p + q$ is even.

As the array is scanned away from broadside, each grating lobe (in directional cosine space) will move a distance equal to the sine of the angle of scan and in a direction determined by the plane of scan. To ensure that no grating lobes enter visible space (real space), the element spacing must be chosen so that for the maximum scan angle θ_m the movement of a grating lobe by $\sin \theta_m$ does not bring the grating lobe into visible space. If a scan angle

of 60° from broadside is required for every plane of scan, no grating lobe may exist within a circle of radius $1 + \sin \theta_m = 1.866$. The square lattice that meets this requirement has

$$\frac{\lambda}{d_x} = \frac{\lambda}{d_y} = 1.866 \quad \text{or} \quad d_x = d_y = 0.536\lambda \quad (20-23)$$

Here, the area per element is

$$dx \cdot dy = (0.536\lambda)^2 = 0.287\lambda^2 \quad (20-24)$$

For an equilateral triangular lattice, the requirement is satisfied by

$$\frac{\lambda}{d_y} = \frac{\lambda}{\sqrt{3}d_x} = 1.866 \quad \text{or} \quad d_y = 0.536\lambda, d_x = 0.309\lambda \quad (20-25)$$

If we recall that elements are located only at every other value of mn , the area per element is

$$2d_x d_y = 2(0.536\lambda)(0.309\lambda) = 0.322\lambda^2 \quad (20-26)$$

For the same amount of grating lobe suppression, the saving in the number of elements for the triangular versus the rectangular lattice is 15.5 percent. The grating lobe contour plots illustrating the previous example are shown in Figure 20-9.

Directivity of Linear Arrays

Array directivity is defined as the ratio of power density per unit of solid angle at the peak of the main beam to the average power radiated per unit solid angle over all space.

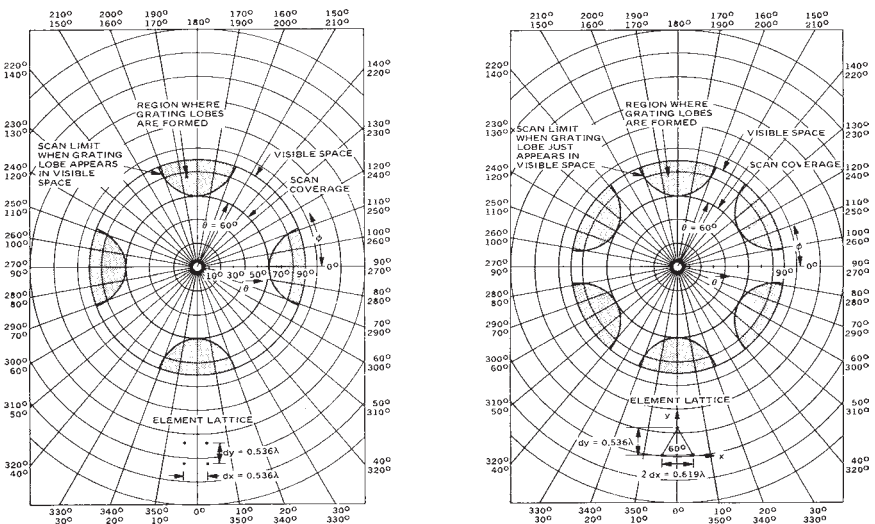


FIGURE 20-9 Grating lobe contours for rectangular and triangular lattices

From this definition, it follows that the directivity of the main beam at angle θ_0 is given by

$$D(\theta_0) = \frac{|E(\theta_0)|^2}{\frac{1}{4\pi} \int_{\text{all space}} |E(\theta_0)|^2 d\Omega} \quad (20-27)$$

where $d\Omega = \sin \theta d\theta d\phi$. For a linear array of isotropic radiators arranged as shown in Figure 20-10, $E(\theta)$ is independent of ϕ . Therefore, $\int |E|^2 d\Omega = 2\pi \int |E|^2 \cos \theta d\theta$. Using $v = \sin \theta dv = \cos \theta d\theta$, we have

$$D(v_0) = \frac{2 |E(v_0)|^2}{\int_{-1}^{+1} |E(v)|^2 dv} \quad (20-28)$$

Using Eq. 20-8 and integrating yield

$$D(v_0) = \frac{\left| \sum_{n=0}^{N-1} a_n \right|^2}{\sum_{n=0}^{N-1} \sum_{m=0}^{N-1} a_m a_n \frac{\sin 2\pi(d/\lambda)(n-m)}{2\pi(d/\lambda)(n-m)}} \quad (20-29)$$

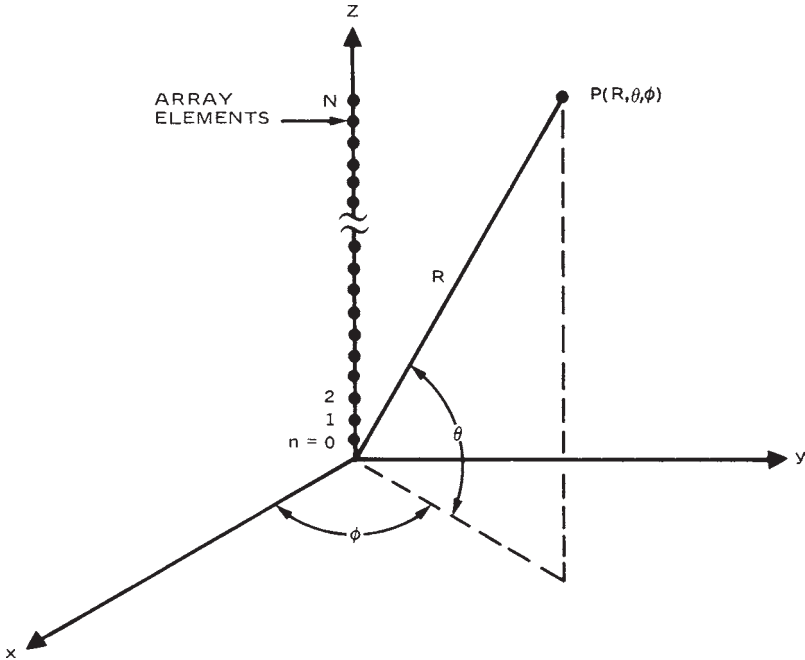


FIGURE 20-10 *N*-element linear array geometry

The denominator of Eq. 20-29 prevents a simple analysis of the directivity. When all excitation coefficients are equal, the directivity simplifies to

$$D(v_0) = \frac{N^2}{N + 2 \sum_{n=1}^{N-1} \frac{(N-n) \sin 2nkd}{2nkd}} \quad (20-30)$$

This result is plotted in Figure 20-11 for arrays of 2 to 24 elements, with spacings up to 1.2λ . It may be noted that the directivity drops abruptly at the appearance of the first grating lobe.

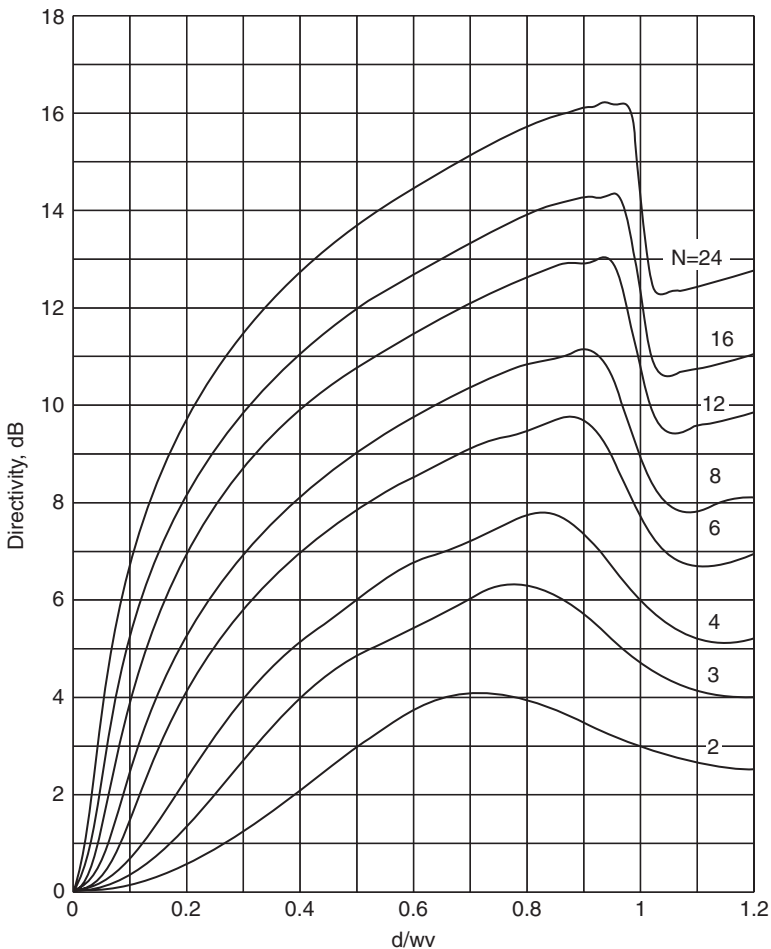


FIGURE 20-11 Directivity of linear array of isotropes (after Hansen³ © Wiley 1998)

Directivity of Planar Arrays

The directivity of a planar array is defined as in Eq. 20-31 with appropriate notation changes to include both spatial angles:

$$D(\theta_0, \phi_0) = \frac{|E_a(\theta_0, \phi_0)|^2}{\frac{1}{4\pi} \int_{\text{all space}} |E_a(\theta, \phi)|^2 d\Omega} \quad (20-31)$$

where $d\Omega = d\theta d\phi \sin\theta$.

Even for an array of isotropic elements, equally excited, Eq. 20-31 is difficult to evaluate as the denominator contains the square of the array series. Directivity in terms of $\sin x/x$ functions of complicated arguments is given in Hansen.³ Fortunately, for scan angles not near endfire, the directivity is given approximately by

$$D(\theta_0, \phi_0) = \frac{4\pi A}{\lambda^2} \eta \cos\theta_0 \quad (20-32)$$

where η is the excitation or taper efficiency. Near endfire the beam shape and directivity change appreciably. For a large array

$$D(\theta_0, \phi_0) = 4\pi \eta N \frac{d_x d_y}{\lambda^2} \cos\theta_0 \quad (20-33)$$

The directivity expressions as shown in Eqs. 20-32 and 20-33 are for an array aperture which is impedance matched for all beam scan angles; for example, the aperture is transparent for a plane wave incidence at any scan angle θ_0 without reflections. However, in practice, the aperture is mismatched at some scan angles. Therefore, the array gain of the antenna is given approximately by

$$G(\theta_0, \phi_0) = \frac{4\pi A}{\lambda^2} \eta \cos\theta_0 [1 - |\Gamma(\theta_0, \phi_0)|^2] \quad (20-34)$$

where $|\Gamma(\theta_0, \phi_0)|$ is the amplitude of the array aperture reflection coefficient at the scan angle θ_0, ϕ_0 .

In general, the antenna beamforming network behind the array aperture has ohmic losses, such as the losses in the phasers, power combiners, etc.

To obtain an accurate gain formula, the effects of mutual impedance must be included. The resulting array gain versus scan angles is called Scan Element Pattern; see Section 20.4 for more details.

20.3 PATTERN CONTROL

Pattern Control of Linear Arrays

The far-field pattern properties of most frequent concern to the array designer are the array sidelobe level, array gain, and the beamwidth. All these properties depend upon the amplitude taper applied to the excitation coefficients a_n 's of Eq. 20-5. In general, a stronger amplitude taper across the array aperture has the effect of reducing the sidelobes at the expense of increased beamwidth and reduced aperture efficiency (antenna gain).

In the days before computers, aperture distributions were chosen for easy integrability: for example, cosine-on-a-pedestal. In the 1950s, Taylor at Hughes Aircraft developed aperture distributions based on physics principles. These principles required symmetric distributions, pedestals (so far-out sidelobes decayed as $1/u$), far-out zeroes separated by unity, and zeroes should be located on the w unit circle (Taylor^{3,7}). These widely used distributions include two for linear and rectilinear arrays, and two for circular arrays.

Taylor \bar{n} Distribution

The Taylor \bar{n} distribution⁷ provides a certain number of equal sidelobes, symmetrically located on both sides of the main beam with the amplitudes of the remaining sidelobes decreasing monotonically. For a given design sidelobe level, the Taylor distribution provides narrowest beamwidth. The Taylor \bar{n} amplitude distribution is of the form

$$A(x, A, \bar{n}) = \frac{1}{2\pi} \left\{ F(0, a, \bar{n}) + 2 \sum_{n=1}^{\bar{n}-1} F(n, A, \bar{n}) \cos \frac{n\pi x}{L} \right\} \quad (20-35)$$

$$F(n, A, \bar{n}) = \frac{[(n-1)!]^2 \prod_{m=1}^{\bar{n}-1} \left(1 - \frac{n^2}{\sigma^2 [A^2 + (-\frac{1}{2})^2]} \right)}{(\bar{n}-1+n)! (\bar{n}-1-n)!}$$

where

x = distance from center of aperture

L = total length of aperture

$A = 1/\pi \text{ arc cosh } R$

R = design sidelobe voltage ratio

$\sigma = \frac{\bar{n}}{\sqrt{A^2 + (\bar{n} + \frac{1}{2})^2}}$

\bar{n} = number of equiamplitude sidelobes adjacent to main beam on one side

Corresponding to this amplitude distribution, the beamwidth of a one-wavelength source is given by

$$\beta = \sigma \beta_0 \quad (20-36)$$

where

$$\beta_0 = \frac{2}{\pi} \sqrt{(\text{arc cosh } R)^2 - \left(\text{arc cosh } \frac{R}{\sqrt{2}} \right)^2} \quad (20-37)$$

For an array with aperture length L , the beamwidth is given by

$$\text{Beamwidth} = \sigma \beta_0 \frac{\lambda}{L} \text{ rad} \quad (20-38)$$

Beamwidth as a function of the design sidelobe level for various values of \bar{n} is given in Table 20-1.

TABLE 20-1 Design Sidelobe Level and Beamwidth for Taylor Distributions

Design sidelobe level, dB	R (sidelobe voltage ratio)	$\frac{180\beta_0}{\pi}$ (degrees)	Values of the parameter σ									
			A^2	$\bar{n} = 2$	$\bar{n} = 3$	$\bar{n} = 4$	$\bar{n} = 5$	$\bar{n} = 6$	$\bar{n} = 7$	$\bar{n} = 8$	$\bar{n} = 9$	$\bar{n} = 10$
15	5.62341	45.93	0.58950	1.18689	1.14712	1.11631	1.09528	1.08043	1.06969	1.06112	1.05453	1.04921
20	10.00000	51.17	0.90777	1.12549	1.12133	1.10273	1.08701	1.07490	1.06554	1.05816	1.05223	1.04738
25	17.7828	56.04	1.29177		1.09241	1.08598	1.07728	1.06834	1.06083	1.05463	1.04948	1.04518
30	31.6228	60.55	1.74229			1.06934	1.06619	1.06079	1.05538	1.05052	1.04628	1.04262
35	56.2341	64.78	2.25976				1.05386	1.05231	1.04923	1.04587	1.04264	1.03970
40	100.0000	68.76	2.84428					1.04298	1.04241	1.04068	1.03808	1.03643

Taylor One-Parameter Distribution

The modified $\sin \pi u/\pi u$ taper of the Taylor One-Parameter distribution³ produces monotonically decreasing sidelobes. The radiation pattern has a main lobe of adjustable amplitude and a sidelobe structure similar to that of the radiation from a uniformly illuminated source. The amplitude distribution is of the form

$$a(x) = I_0 \left[\pi \beta \sqrt{1 - \left(\frac{2x}{L} \right)^2} \right] \quad (20-39)$$

where

x = distance measured from center of aperture

L = total length of aperture

I_0 = zero order Bessel function of third kind

β = parameter fixing ratio R of main beam amplitude to amplitude of first sidelobe by
 $R = 4.60333 \sinh \pi \beta / \pi \beta$.

The array pattern corresponding to the previous amplitude distribution is given by

$$E_a(u) = \frac{\sin \pi \sqrt{u^2 - B^2}}{\pi \sqrt{u^2 - B^2}} \quad u \geq B \quad (20-40)$$

$$E_n(u) = \frac{\sinh \pi \sqrt{B^2 - u^2}}{\pi \sqrt{B^2 - u^2}} \quad u \leq B \quad (20-41)$$

where

$$u = (L/\lambda) \sin \theta \quad (20-42)$$

The beamwidth and aperture efficiency as a function of the design sidelobe ratio are shown in Table 20-2. Figure 20-12 shows a Taylor One-Parameter pattern for -25 dB sidelobe ratio. For details see Hansen.³

Dolph-Chebyshev Distribution

In 1946 Dolph⁸ showed that a linear aperture (array) would have maximum directivity for a given sidelobe level if all sidelobes were of equal height. This was accomplished by effecting

TABLE 20-2 Design Sidelobe Level, Beamwidth, and Aperture Efficiency for Modified $\sin \pi u/\pi u$ Distributions

Sidelobe ratio, dB	B	$180\beta/\pi$, beamwidth of a one-wavelength source, °	η , aperture efficiency
13.26	0	50.75	1.0000
15	0.355769	52.884	0.9931
20	0.738600	58.659	0.9330
25	1.02292	63.938	0.8626
30	1.27616	68.775	0.8014
35	1.51363	73.232	0.7509
40	1.74148	77.378	0.7090

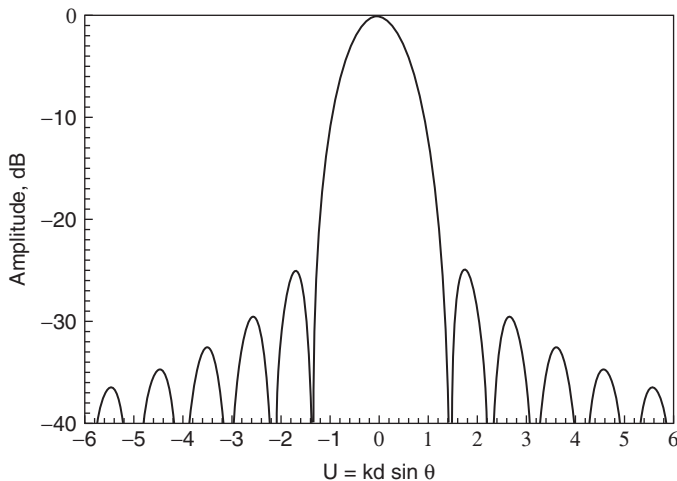


FIGURE 20-12 Taylor one-parameter pattern, SLR = 25 dB

a correspondence between the array polynomial and the Chebyshev polynomial. These distributions are mainly of academic interest for several reasons: the far-out sidelobes are high; for some cases the aperture distribution is non-monotonic; and the aperture efficiency is low.³

Bayliss Difference Pattern

For monopulse radar or angle-of-arrival systems, a difference pattern is useful. It typically has a split beam with zero gain at broadside. The steep slope of the pattern on either side of broadside allows accurate angle determination. The simplest arrangement uses a uniformly excited linear array with one half 180° out of phase with the other half. Unfortunately the highest sidelobe is -10.6 dB, a level often unacceptable. A remedy was provided by Bayliss,⁹ who developed a difference pattern based on the Taylor \bar{n} distribution discussed earlier. This allows lower difference pattern sidelobes to be selected versus excitation efficiency. Figure 20-13 shows a typical Bayliss pattern with -25 dB sidelobe level, while Figure 20-14 gives the amplitude distribution. Details of the equations and efficiencies are given in Hansen.³

Pattern Control of Planar Arrays

If markedly different patterns are desired in the two orthogonal planes ($\cos \alpha_x$ and $\cos \alpha_y$ planes), it is often advantageous to use rectangularly-shaped arrays with sides parallel to the specified planes and separable illumination; that is, $a_{mn} = a_m a_n$, so that

$$\begin{aligned}
 E_a(\cos \alpha_x, \cos \alpha_y) &= \sum_m \sum_n a_{mn} e^{jk[md_x(\cos \alpha_x - \cos \alpha_{x0}) + nd_y(\cos \alpha_y - \cos \alpha_{y0})]} \\
 &= \sum_m a_m e^{jkm d_x(\cos \alpha_x - \cos \alpha_{x0})} \sum_n a_n e^{jkn d_y(\cos \alpha_y - \cos \alpha_{y0})} \\
 &= E_{a1}(\cos \alpha_x) \cdot E_{a2}(\cos \alpha_y)
 \end{aligned} \tag{20-43}$$

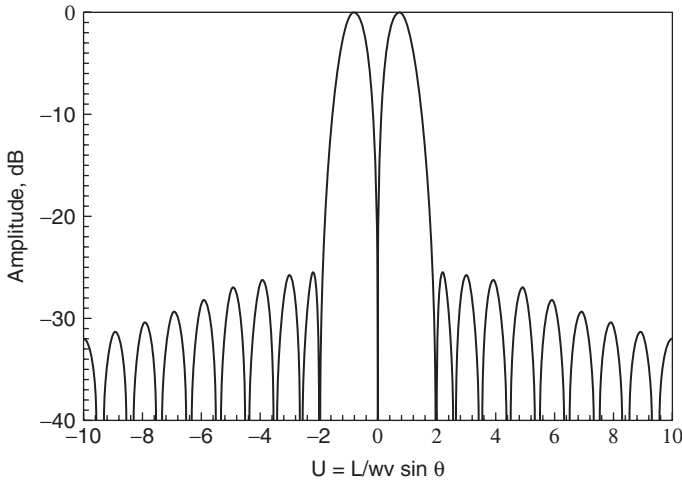


FIGURE 20-13 Bayliss difference pattern, $\bar{n} = 5$

where E_{a1} and E_{a2} are synthesized by linear array techniques. In this case, the amplitude distributions described in the subsection “Pattern Control of Linear Arrays” can be applied along each orthogonal axis of the planar array to achieve the desired sidelobe control. One special feature of this separable illumination is that the sidelobe levels in the $\cos \alpha_x$ plane are those of $E_{a1}(\cos \alpha_x)$ alone, since E_{a2} is a maximum. Away from both the $\cos \alpha_x$ and the $\cos \alpha_y$ axes (diagonal planes), the sidelobes are lower as a result of the product of E_{a1} and E_{a2} , both less than their maxima. For example, the array factor for a uniformly illuminated rectangular aperture was given by Eq. 20-18. The array factor yields first sidelobes of about -13.2 dB below the main beam along the two orthogonal axes. Off the axes, the sidelobes are lower. In particular, along the 45° diagonals, the first sidelobes are twice as low; -26.4 dB.

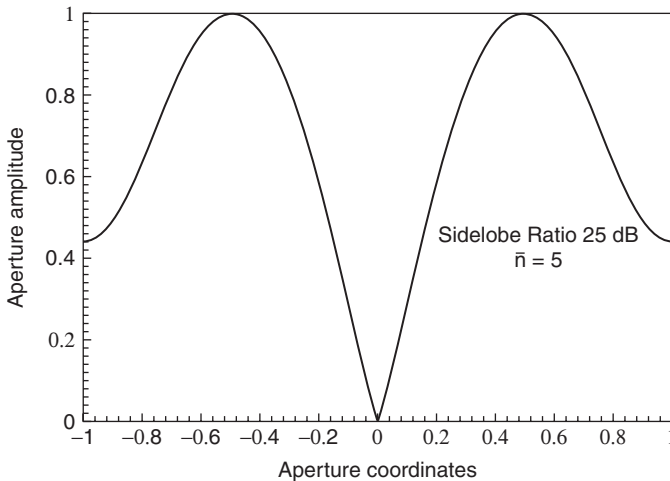


FIGURE 20-14 Bayliss difference pattern amplitude

Consequently, for a given peak sidelobe level, the taper efficiency ($\eta = \eta_1 \eta_2$) is lower for a separable illumination than that which can be realized by nonseparable means, although the average sidelobe level should be comparable to tapers of like values of η .

If a circularly symmetric pattern is desired, an array shape approximating a circle is preferred. Here, the array symmetry refers to only the symmetry of the beam shape and near-in-sidelobe structure. The element grid will still govern the grating lobe location. In most cases, the amplitude distributions for circular arrays are nonseparable.

Taylor Circular n -bar Distribution

This low sidelobe distribution^{10,11} produces a pattern that is rotationally symmetric. Design principles are similar to those of the linear \bar{n} distribution.³ Useful data are given in Table 20-3.

Hansen Circular One-Parameter Distribution

This symmetric distribution¹² uses a modified $I_1(\pi u)/\pi u$ in the higher part of the main beam, and a modified $J_1(\pi u)/\pi u$ for the rest of the main beam and sidelobes. The Bessel arguments are replaced by $\pi\sqrt{u^2 - H^2}$, where H corresponds to the B constant of the linear One-Parameter distribution. Table 20-4 gives design data.³

For the special case of a uniformly illuminated circular array, corresponding to $H = 0$ in the One-Parameter distribution, the array pattern is given by

$$E_a(\theta) \approx \frac{J_1(ka \sin \theta)}{ka \sin \theta} \quad (20-44)$$

where $J_1(x)$ is the first-order Bessel function. The function has an HPBW coefficient of 58.96° , defined by

$$\text{Beamwidth} = 58.96 \frac{\lambda}{2a} \quad (20-45)$$

where a is the radius of the circular aperture. The first sidelobes are approximately -17.5 dB.

There is also a circular Bayliss difference pattern. Details and references are given in Hansen.³

20.4 RADIATING ELEMENT DESIGN AND MUTUAL COUPLING

The discussion on array theory in Section 20.2 is based on elements with isotropic patterns. In practice, however, the radiation patterns of real elements are nonisotropic, and the impedance of the elements varies as a function of scan caused by the mutual coupling between the elements. In fact, the pattern of an element in the environment of an array is markedly different from the pattern of an isolated element in amplitude, phase, and, perhaps, polarization as well.

One of the functions performed by the antenna is to provide a good match between the radar transmitter and free space. This means that the driving point impedance looking into the element in the array environment must be matched to their generator impedance.

TABLE 20-3 Design Sidelobe and Beamwidth of Circular Taylor \bar{n} Distributions

Design sidelobe level, dB	η (sidelobe voltage ratio)	Ideal beamwidth β_o	Values of the parameter σ									
			A^2	$\bar{n} = 3$	$\bar{n} = 4$	$\bar{n} = 5$	$\bar{n} = 6$	$\bar{n} = 7$	$\bar{n} = 8$	$\bar{n} = 9$	$\bar{n} = 10$	
15	5.62341	45.93	0.58950	1.2382	1.1836	1.1485	1.1244	1.1069	1.0937	1.0833	1.0751	
20	10.00000	51.17	0.90777	1.2104	1.1692	1.1398	1.1186	1.1028	1.0906	1.0810	1.0732	
25	17.7828	56.04	1.29177	1.1792	1.1525	1.1296	1.1118	1.0979	1.0870	1.0782	1.0709	
30	31.6228	60.55	1.74229	1.1455	1.1338	1.1180	1.1039	1.0923	1.0827	1.0749	1.0683	
35	56.2341	64.78	2.25976		1.1134	1.1050	1.0951	1.0859	1.0779	1.0711	1.0653	
40	100.0000	68.76	2.84428		1.0916	1.0910	1.0854	1.0789	1.0726	1.0670	1.0620	

TABLE 20-4 Design Sidelobe Level, Beamwidth, and Aperture Efficiency for Hansen Circular One-Parameter Distribution

Sidelobe ratio, dB	H	$180\beta/\pi$, beamwidth of a one-wavelength source	η , aperture efficiency
17.57	0	58.96	1.000
20	.4872	61.80	.9786
25	.8899	67.25	.8711
30	1.1977	72.24	.7595
35	1.4708	76.79	.6683
40	1.7254	81.02	.5964

If the antenna aperture is not matched to free space, power will be reflected back toward the generator, resulting in a loss in radiated power. In addition, a mismatch produces standing waves on the feed line to the antenna. Therefore, while the antenna is radiating less power, individual components must be designed to handle more peak power. In a scanning array, the impedance of a radiating element varies as the array is scanned, and the matched problem is considerably more complicated. In some instances, spurious lobes may appear in the array pattern as a consequence of the mismatch. Furthermore, there are conditions in which an antenna that is well matched at broadside may have some angle of scan at which most of the power is reflected (scan blindness).

Design Criteria for Elements

The most commonly used elements for phased arrays are dipoles,¹³ waveguide slots,¹⁴ open-ended waveguides,¹⁵ TEM horns,^{16,17} spirals,¹³ and microstrip patch elements.¹⁸ The TEM horn is sometimes called the Vivaldi antenna, a whimsical name that has no content and is deprecated. The selection of an element for a particular application must be based upon the following considerations:

1. The required area of the element is small enough to fit within the allowable element spacing and lattice without the formation of grating lobes. In general, this limits the element to an area of a little more than $\lambda^2/4$.
2. The Scan Element Pattern of the element provides the appropriate aperture matching over the required scan coverage (see Section 20.4).
3. The polarization and power handling capability (both peak and average power) meet the system requirements.
4. The physical construction of the element must be able to withstand environmental requirements such as thermal, shock, and vibration requirements.
5. The element must also be inexpensive, reliable, and repeatable from unit to unit since many hundreds or thousands of elements are required in a large phased-array antenna.

Since the impedance and the pattern of an element in an array are determined predominantly by the array geometry and environment, the radiator may be chosen to best suit the feed system and the physical requirements of the antenna. For example, if the element is fed from a stripline or microstrip phaser, a stripline or microstrip patch would be a logical choice for that element.

After selecting the proper type of element on the basis of the criteria as stated previously, the next step is to optimize the performance by following the procedure outlined next:

1. Select an element configuration and spacing that do not generate grating lobes.
2. Formulate an analytical model of the radiating element in the array aperture, and optimize the performance by varying the design parameters of the element such as the spacing between the elements, the aperture size, or the distance of the element above ground. The Scan Impedance may be modeled by Floquet's infinite array theory, or by numerical methods such as the Moment Method, Finite Elements, or FDTD.
3. Using the analytical results from Step 2, fabricate waveguide simulators to measure the Scan Impedance at discrete scan. The design parameters of the element may have to be modified in order to optimize element performance. This measurement also verifies the calculated results of Step 2.
4. Fabricate a small test array and measure the Scan Element Pattern using the array in receive mode, over the required frequency band. The Scan Element Pattern describes the variation in array gain as a function of beam scan angle. Large aperture mismatches at any scan angle or scan plane will show up as dips in the Scan Element Pattern.
5. The Scan Element Pattern measurement of Step 4 in conjunction with the Scan Impedance measurement of Step 3 establishes the performance characteristics of the final element design.

Scan Element Pattern (SEP)

Scan Element Pattern, formerly called Active Element Pattern, is gain per element versus scan angles. *Active* is deprecated because it has electron device connotations. Mutual impedance effects are included. For dipole type radiating elements, drive voltages are applied; the resulting currents are used to calculate SEP.

$$SEP = \frac{120 F^2(\theta, \phi) \sum_{n=1}^N I_n \sum_{n=1}^N I_n^*}{N \sum_{n=1}^N I_n I_n^* R_n} \quad (20-46)$$

Here the element pattern is F and R_n is the Scan Resistance of the n th element. Note that summing fields from one element excited at a time is incorrect; the correct set of currents requires that all elements be excited.¹⁹ Measurement of SEP thus requires the array to be used as a receiver, so that all element amplitudes and phases are correct. This is easily done in a gain measurement setup, using a compact range, or an anechoic chamber. SEP is also usually validated by a computer-based simulation code, or by waveguide simulators; see below. Scan Impedance (SI), formerly called Active Impedance, is not directly measurable. Network analyzer setups excite one array element only. A method that can be used is the Load Pull method (Van Wagoner and Hansen²⁰), where in a receiving array the matching impedance for one element is adjusted for maximum received power. The SI is then the complex conjugate of this impedance. Scan Impedance, which is a function of scan angles θ , ϕ and frequencies, is important as it affects the impedance match and power transfer. Scan Element Pattern, also a function of scan and frequency, tells the radar or communications systems designer how well the system works. Note that SEP is not an antenna "pattern"; it shows gain versus scan. Figure 20-15 shows SEP for a dipole array; Figure 20-16 is for the array with a ground plane.

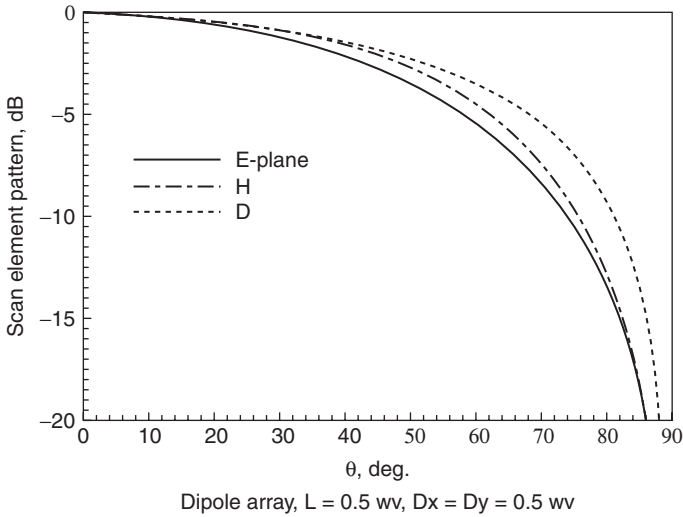


FIGURE 20-15 Scan element pattern, dipole array

Waveguide Simulators

Scan Impedance of an element in an infinite array can be simulated^{21,22} by placing an element, or sub-array, in a rectangular waveguide. The plane waves that compose the dominant waveguide mode simulate the array at the plane wave angle. The Scan Impedance is obtained from the reflection coefficient in the guide. Figure 20-17 shows a typical simulator in rectangular waveguide. Parts of adjacent elements are included in the unit cell. This infinite array model can predict with good accuracy the array impedance and impedance

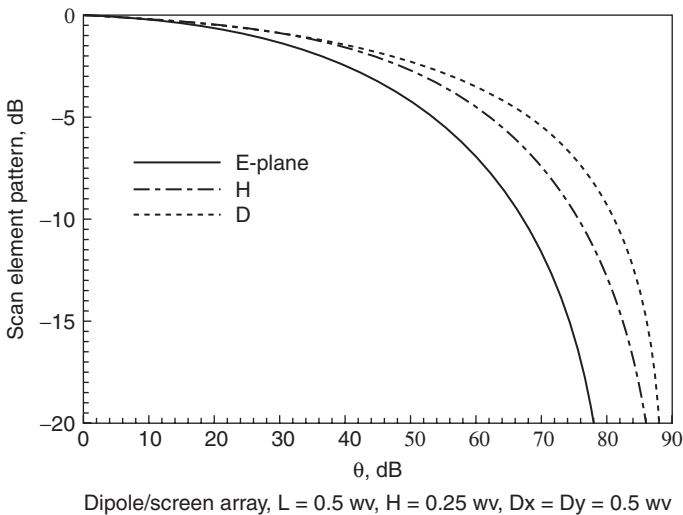


FIGURE 20-16 Scan element pattern, dipole array with screen

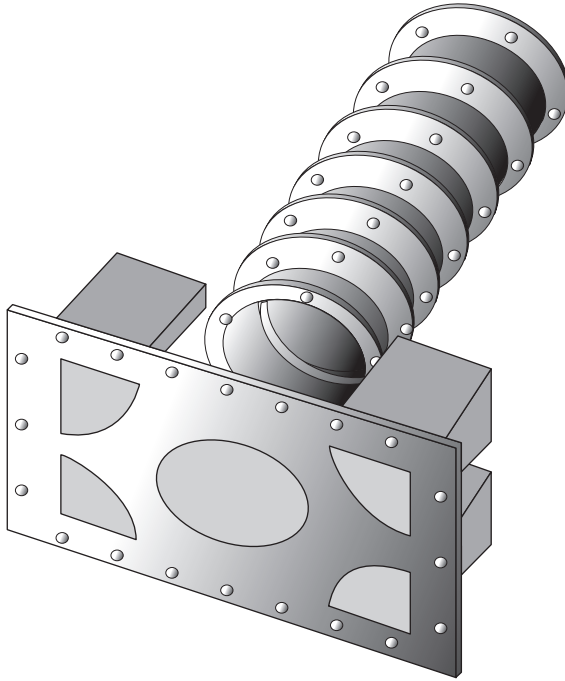


FIGURE 20-17 Waveguide simulator for hexagonal lattice (after Hannan and Balfour²¹ © IEEE 1972)

variation with scan for a finite array. Even arrays of modest proportions (less than 1000 elements) have been in reasonable agreement with the results predicted for an infinite array. The reason for this is that, in practice, the mutual coupling between the elements decays as $1/x^2$, where x is the distance between the elements. Hence, the effect of mutual coupling is quite localized in cases of practical interest. All the elements with the exception of the edge elements in the array behave as though they are in an infinite array environment. For large arrays, the effect of edge elements on array performance is negligible.

On all arrays, the edge elements experience impedance changes. These effects have been modeled by a Gibbs type phenomenon.²³ Edge effects are somewhat larger for E-plane scan.^{24,25}

20.5 FEED DESIGN FOR PHASED ARRAYS

This section presents a survey of the various methods of feeding the array elements to form the required beam patterns in space. For simplicity, linear array configurations will be used as illustrations. Extrapolation to a planar array can be accomplished either with a single technique or by using a combination of two different techniques; for example, one technique for rows and the other for column. In general, most of the beamforming feeds for phased arrays can be categorized into three basic groups: (1) networks, (2) multiple beams, and (3) lenses. A description of the various types of feeds in each of these groups is given next.

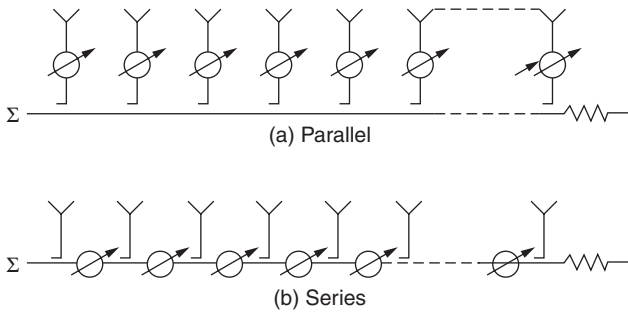


FIGURE 20-18 End-fed series feeds

Network Feeds

The various types of network feeds can be classified into two groups: series feeds and parallel feeds. In each group, the feeds can be designed to provide either a single pencil beam or monopulse sum-and-difference beams. A description of the various types of feeds in each group follows.

Series Feeds

Figure 20-18 illustrates several types of series feeds. In all cases the path length to each radiating element must be computed as a function of frequency and taken into account when setting the phaser. Figure 20-18 illustrates two examples of end-fed serial feed systems in which the elements are arranged serially along the main line. They consist of a main transmission line from which energy is tapped (in the case of transmission) through loosely coupled junctions to feed the radiating elements. To steer the beam, phasers are added in either of the branch lines feeding the radiating elements, as shown in Figure 20-18a, or in the main line, as shown in Figure 20-18b. The amplitude taper is established by properly designing the couplers at the junctions. For example, if all the junction couplers are identical, the amplitude taper envelope will be approximately exponential.

Mechanical simplicity is, perhaps, the greatest advantage of these two configurations over others to be discussed. They are easy to assemble and to construct. The configurations are easily adapted to construction in waveguides, using cross guide directional couplers as junctions. They are potentially capable of handling full waveguide power at the input (within the limitations of the phaser). However, they suffer losses associated with the corresponding length of a waveguide plus that of the phasers. Perhaps the most severe limitations of this feed system is its dependence of pointing angle on frequency.

The configuration of Figure 20-18a places lower power handling demands on the phasers and also results in a lower system loss for a given phaser loss. On the other hand, the configuration of Figure 20-18b has an advantage in that all phasers will have identical phase shifts for a given pointing angle—a property which results in simplified array control. However, the total system loss of Figure 20-18b is higher than that of Figure 20-18a.

Monopulse sum-and-difference beams can be formed by feeding the array in the middle instead of from the end (see Figure 20-19a). This approach, however, would not be able to provide good sum-and-difference patterns simultaneously. For example, if the couplers in the feed are designed to provide low sidelobes for the sum pattern, the resultant pattern for

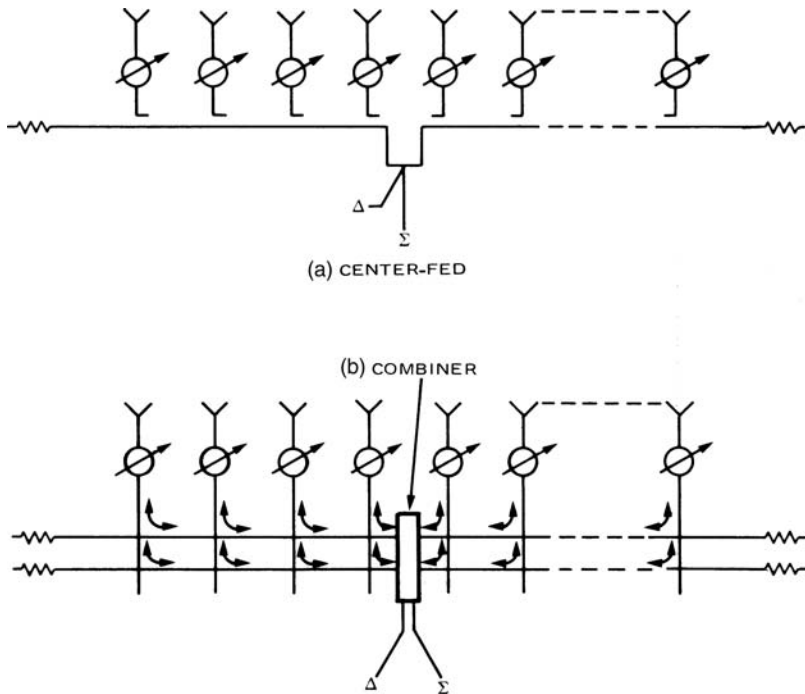


FIGURE 20-19 Center-fed series feeds with monopulse sum-and-difference beams

the difference beam would have high sidelobes. At the cost of some additional complexity, this sidelobe problem can be solved by the method shown in Figure 20-19b. Two separate center-feed feed lines are used and combined in a network to give sum-and-difference-pattern outputs.²⁶ Independent control of the two amplitude distributions is possible. For efficient operation, the two feed lines require distributions that are orthogonal to each other; that is, the peak of the pattern of one feed line coincides with the null from the other. The aperture distributions are respectively even and odd.

The bandwidth of a series feed can be increased by making the path length from the input to each output of the branch lines all equal, as shown in Figure 20-20. However, if the bandwidth is already limited by the phasers and the couplers, very little benefit can be derived from this approach at the cost of a considerable increase in size and weight. The network of Figure 20-20 simplifies the beam steering computation since the correction for path length differences is no longer necessary.

The frequency dependence associated with series feeds can be reduced by the use of parallel feeds at a cost of slightly more complex mechanical structure. The frequency dependence (or change in beam pointing angle with frequency) of parallel feed configurations depends mainly on the frequency characteristics of the phasers and the couplers at the junctions of the feed, if the line lengths from the transmitter to the elements are all made equal. Thus, if variable time-delay phasers were used in place of the constant phase type of phasers at each element, the beam-pointing angle would be essentially independent of frequency. The frequency dependence of a parallel feed using constant-phased-type phasers will be discussed later.

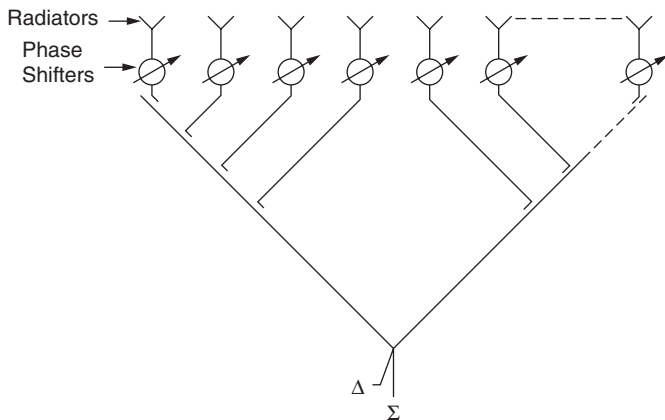


FIGURE 20-20 Equal-path length feed

The most common technique for realizing parallel feeds is the *matched-corporate-feed structure* of Figure 20-21. In this configuration, the corporate feed is assembled from matched hybrids (four-port junction devices with the isolated ports terminated into matched loads). The out-of-phase components of mismatch reflections from the aperture and of other unbalanced reflections are absorbed in the terminations. The in-phase and balanced components are returned to the input, and power reflected from the aperture is not reradiated. When nonreciprocal phasers (the two-way path length through the phaser is a constant independent of the phaser setting) are used in place of reciprocal phasers at each radiating element, the couplers at the junctions of the corporate feed can be reactive Tee's (three-port devices) instead of matched hybrids. In this case, however, the reflections from the aperture will add up in phase at the input to the feed. Therefore, a high power isolator must be used to protect the transmitter and/or a high power limiter be used to protect the receiver.

A method of achieving sum-and-difference monopulse beams for a parallel feed is shown in Figure 20-22. The signals from a pair of elements, which are located symmetrically opposite from the centerline of the array, are combined in a magic *T* (matched hybrid) to form their sum-and-difference signals. The sum signals from all the pairs across the array aperture are then combined in a power combiner network to form a sum beam. The desired amplitude distribution for the sum beam can be achieved by proper weighting of the signals in the power combiner network. The difference signals from all the pairs are combined in a separate power combiner network to form a difference beam. Amplitude weighting in the

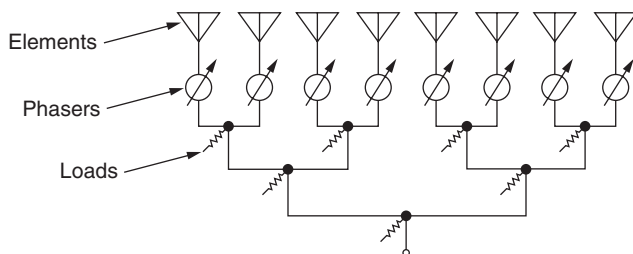


FIGURE 20-21 Corporate feed

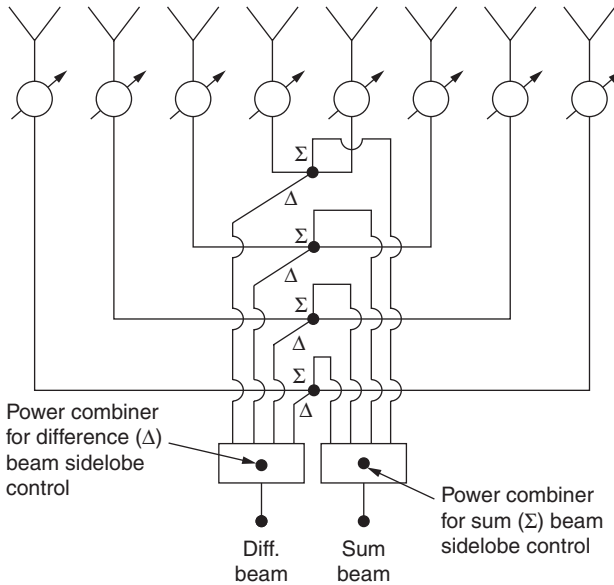


FIGURE 20-22 Constrained monopulse feed with independently controlled sum-and-difference sidelobes

two combiner networks can be made differently to obtain low sidelobes for both the sum and the difference beams. For example, the low sidelobe amplitude distribution for a sum beam is in the form of a Taylor distribution, whereas a Bayliss distribution is used for low sidelobe difference beams.⁹ By using the method, independent control of both the sum and the difference beams is achieved.

Phasers and Time Delay

A key component of many phased arrays is the phaser (or phase shifter). There are many phaser types and configurations; these are the subjects of Chapter 21. With the current interest in wideband antennas, time delay units are appropriate. (Note that all time delay is “true.”) In principle, a digital delayer could be placed at each array element, with the maximum delay of $(N - 1)kd \sin \theta$, where the (linear) array has N elements with spacing d , and scan angle θ . For a large array, each delayer must have many bits of delay. In practice a phaser with three to five bits is located at each element, with delayers located down the feed structure. A simplistic approach of installing the smallest delay bit at the element combiners, then the next larger delay bit at the next layer of combiners, and so on, is not recommended due to the generation of high quantization lobes at some frequencies.²⁷ For wideband systems, a good arrangement locates a delay bit and phaser in one arm of a binary combiner; larger delay bits and phasers are placed in one arm of each successive combiner. For any arrangement of delayers and phasers, extensive computer simulations are essential to insure against unacceptable QL and main beam splitting.

Each system will have 1σ amplitude and phase tolerances required to provide the desired array sidelobe and gain performance. With long arrays, a 5 or 10 deg 1σ phase tolerance may be difficult to maintain in a maximum delay of $(L/\lambda) \sin \theta$ deg., or $t_{ns} = 3.33 L \sin \theta$, where array length L is in meters. Delay lines are also lossy, so amplitude

compensation may be necessary for each larger delay bit. Frequency dispersion must always be minimized.

Multiple Beam Feeds

Another class of constrained feeds for phased arrays consists of the multiple beamforming feed network. These feeds can generate multiple simultaneous beams covering a large sector of space. For the same beam-pointing direction, each beam has essentially the gain of a single-beam array of the same size and illumination except for a possible increase in RF circuit losses over a single beam array. A separate beam terminal is provided for each beam of the multiple beamforming feed network.

Butler Matrix

A widely used multiple beam antenna is the Butler matrix.²⁸ As sketched in Figure 20-23, the number of beam ports equals the number of element ports; both are a power of two. The connecting transmission lines are interconnected by 90 degree hybrids. Fixed units of phase provide the proper phase front for each beam. The many unconnected crossovers make implementation more difficult. Note that if the phase units are replaced by equivalent time delay units, the Butler bandwidth decreases.

A signal injected at any of the beam input terminals excites all the radiating elements equally in amplitude, with phase differentials of odd multiples of $180^\circ/N$, where N is the total number of radiating elements or beam terminals. The array patterns of the beam generated are of the form $\sin Nx / \sin x$ since the amplitude taper is uniform. Specifically, they are

$$E_m(v) = \frac{\sin \frac{N}{2} \left[kd - (2m+1) \frac{\pi}{N} \right]}{\sin \frac{1}{2} \left[kd - (2m+1) \frac{\pi}{N} \right]} \quad (20-47)$$

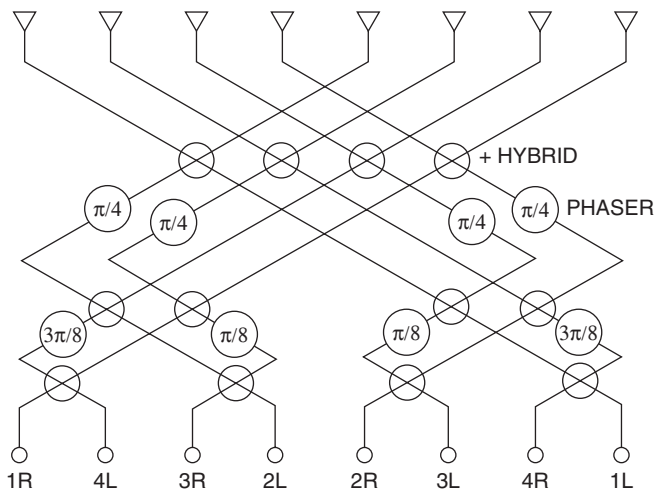


FIGURE 20-23 Butler matrix beamformer (after Hansen³ © Wiley 1998)

with N main beams completely covering one period of v . The beams specified by Eq. 20-47 cross over at a relative amplitude of $2/\pi$ (-3.92 dB), and the peaks of the beams are located at the nulls of the other beams. Since these beams are orthogonal to each other, there is no cross-coupling loss between beams. As expected from Eq. 20-47, as the frequency changes the beam positions change, although the individual beamwidths do not. At twice the design frequency the rosette of beams occupies ± 45 deg. And at half the design frequency the rosette of beams contains only half of the beams; the rest are in “invisible space.” Butler matrices can be stacked and cascaded to form a 2-D rosette of beams. There is also a 2-D Butler-type matrix based on a hexagonal array with hexagonal element lattice.³

Rotman Lens

Another popular multiple beam antenna is the Rotman Lens.²⁹ It consists of a flat 2-D wave guiding medium such as parallel plate waveguide, stripline, or microstrip. A set of beam ports is located on a concave arc, and a set of element ports is located on an opposing concave arc; see Figure 20-24. The arcs are generally of different shapes, and the number of beam ports is typically not equal to the number of element ports. From each element port, a specified length of transmission line connects to the radiating element. There are six basic design parameters, and these may be optimized for a given application. In principle, the Rotman lens can have three perfect foci: one on center line and two symmetrically located. In practice the best design provides minimum error in least squares sense. There is more information on design trades in the references.³⁰

Frequency behavior is the opposite of that of the Butler matrix antenna. For the Rotman lens, the beam positions are fixed by the geometry. As frequency deviates from the design frequency, the individual beams become narrower or wider. Beam crossover levels change accordingly.

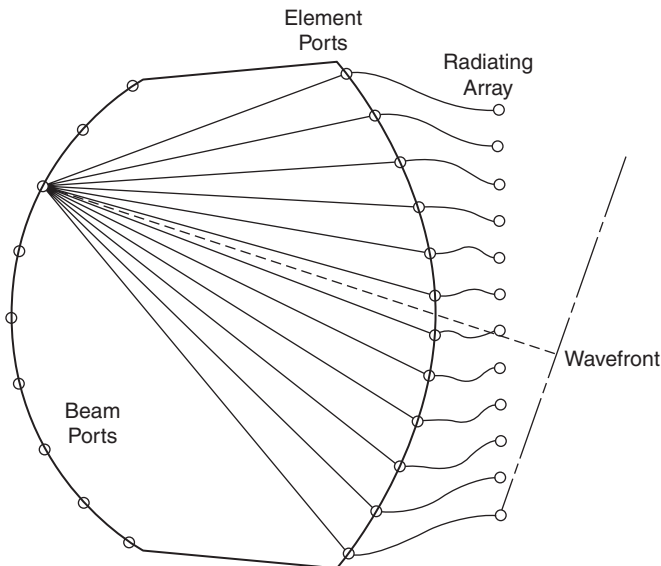


FIGURE 20-24 Rotman lens beamformer (after Hansen³⁰ © IEEE 2002)

Bootlace Lens

Illumination of the array elements in a space-fed antenna is accomplished by optically distributing the source signal (transmitter) through space. This illuminates an array of pickup horns (elements) which are connected to the radiating elements. The advantage of a lens beamforming feed over the network system is simplicity. Disadvantages are a lack of amplitude tapering control and an excessive volume of physical space required to accommodate the feed system. Two basic types of bootlace lens feeds, the transmission type and the reflection type, are shown in Figure 20-25.

In the transmission type, the array elements are the radiating space elements of a feed-through bootlace lens, as shown in Figure 20-25a. These radiating elements are coupled to the pickup elements of the feed-through lens by phasers. Both surfaces of the lens require matching to optimize performance. The pickup surface is illuminated by a primary feed placed at a focal distance f behind the lens. The radiation pattern of the primary feed must be optimized to provide an efficient aperture illumination with little spillover loss. If desired, separate transmit and receive feeds may be used, with the feeds separated by an angle α as shown in Figure 20-25a. The antenna is then rephased between transmitting and receiving so that in both cases the beam points in the same direction. The phasing of the antenna must include a correction for the spherical phase front of the feed. The required phase correction is given by

$$\phi_c = \frac{2\pi}{\lambda} (\sqrt{f^2 + r^2} - f) = \frac{\pi}{\lambda} \frac{r^2}{f} \left[1 - \frac{1}{4} \left(\frac{r}{f} \right)^2 + \dots \right] \quad (20-48)$$

With a sufficiently large focal length, the spherical phase front may be approximated by that of two crossed cylinders, permitting the correction to be applied simply with row and column steering commands. The correction of the spherical phase front can be accomplished by the phasers. Space problems may be encountered in assembling an actual system, especially at high frequencies, since all control circuits must be brought out at the side of

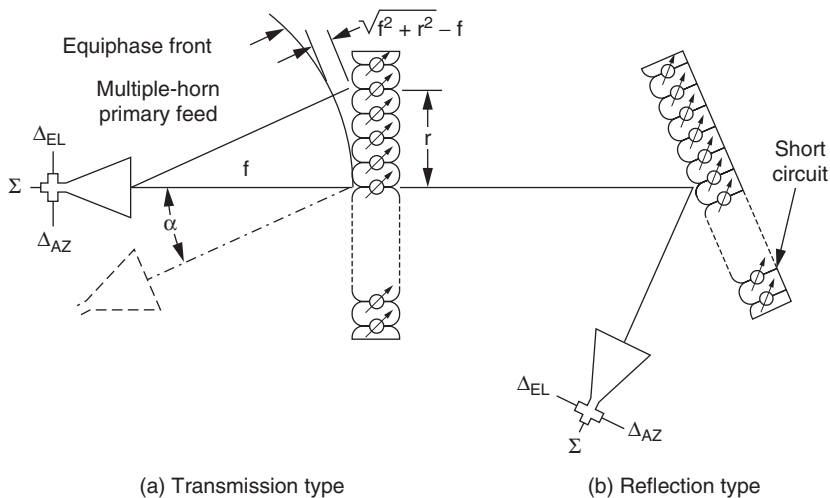


FIGURE 20-25 Bootlace lenses

the aperture. Multiple beams may be generated by adding additional primary feed horns. All the multiple beams will be scanned simultaneously by equal amounts in $\sin \theta$ space. Monopulse sum-and-difference patterns may be generated by taking the sum and difference of two adjacent multiple beams with a magic T or by using multimode feed horns.

In the reflection type, the phased array aperture is used as a reflector, as shown in Figure 20-25*b*. The same radiating element collects and reradiates the signal after it has been reflected from the short circuits terminating the phasers. The phase of the reflected signal is determined by the phaser setting. The phaser must be reciprocal so that there is a net controllable phase shift after the signal passes through the device in both directions. The requirement rules out nonreciprocal phasers. Ample space for phaser control circuits exists behind the reflector. To avoid aperture blocking, the primary feed may be offset as shown. As in the case of a transmission lens, transmit and receive feeds may be separated and the phases separately computed for the two functions. Multiple beams are again possible with additional feeds.

The lens surfaces in Figure 20-25 are flat; if the feed face and the array faces are circular arcs, and if the phasers are replaced by delayers, the lens bandwidth will be significantly improved.

All of the previous information concerns a 1-D bootlace lens with a linear array. Bootlace lenses feeding a planar array are the same in principle. However, a 2-D lens such as the Rotman lens, or a 2-D version of Figure 20-24, can have three perfect focal points. But rotating the 1-D lens does not rotate the foci into circles. A 2-D lens can have up to four foci, but they are generally in one plane, with the result that performances are poorer in the cross plane. Also this requires ellipsoidal faces.³ A 2-D bootlace with spherical cap feed and element port faces will have a modest performance; use of delayers will increase the bandwidth, but off-axis performances will be degraded.

Digital Beamforming

The slow but steady increase in bit-bandwidth product of A/D converters has rekindled interest in digital beamforming for receiving arrays. In such arrays, the output of each element is amplified and filtered as necessary, and then fed to an A/D converter. Or the down converted signal can be sampled, often on the I and Q signals.³¹ Beam scanning, multiple beams, monopulse beams, and sidelobe control are performed digitally. The inter-element delay is

$$\tau = (d/c) \sin \theta_0 \quad (20-49)$$

Where d is the interelement spacing, θ_0 is the beam angle, and c is the velocity of light. Using Nyquist sampling of two samples per cycle, the sampling interval T is

$$T = \lambda / (2c) \quad (20-50)$$

These two times are, in general, not integer multiples of each other. Only certain beam angles allow the beam steering delay values. The most common solution to this dilemma is over sampling. A high rate of over sampling is usually needed to provide many beam positions. Interpolation filtering is followed by summing over a window of time samples and over the array elements.³² An alternate solution interchanges the beamforming and interpolation filtering procedures.

Beamforming can also be performed in the spectral domain, with the advantage that delays become phase multipliers.³³ An FFT is performed on the output of each element, with each output multiplied by $\exp(j\omega\tau)$. Summing yields a beam in frequency. An inverse FFT then provides angular beams in time.

Bandwidth of Phased Arrays

The bandwidth of a phased-array antenna depends upon the types of components, such as elements, phasers, and feed networks, which comprise the array. In practice, most elements for phased arrays are matched over a broad band of frequencies. Therefore, the element design is not a primary factor in the determination of bandwidth. The limitations on bandwidth are determined by the frequency characteristics of the phasers and feed networks.^{34,35} In general, the effects due to the phasers and feed networks are additive, so that if the phaser causes the beam to scan by an amount equal to $\Delta\theta_p$ and the feed causes the beam to scan by $\Delta\theta_f$, then the total beam scan is given by $\Delta\theta_p + \Delta\theta_f$. To establish the bandwidth capability of a phased array, these two effects must be examined. There are more details concerning the bandwidth limitations of phased arrays not covered in this book.³

Phaser Effects

To evaluate the effects caused by the phaser alone, a corporate feed (equal line length parallel feed) is used to illuminate all the radiating elements (see Figure 20-21). This corporate feed exhibits no feed effects since a signal at the input illuminates all the elements with the same phase regardless of frequency. The bandwidth in this case is completely determined by the type of phaser used in the array. In this discussion, consider the effects of two basic types of phasers: (1) time delay phasers and (2) constant phase-type phasers. A detailed description of these two types of phasers is given in Chapter 21.

When delayers are used at each element, the signals received by the elements from an incident wavefront at an angle θ_0 are appropriately time delayed so that they all arrive at the output terminal of the corporate feed at the same time. For example, the amount of time delay at the first element of the array shown in Figure 20-21 is equal to the additional time required for the wavefront to travel to the last element after arriving at the first element. This time delay T , known as aperture fill time, is given by

$$T = \frac{L}{c} \sin\theta_0 \quad (20-51)$$

where

L = total length of array aperture

c = velocity of light

θ_0 = angle of incidence of wavefront from array normal

In this case, the phase distribution across the array aperture produced by the time delay feed matched that of the incident wavefront independent of frequency. Consequently, the beam position remains stationary with frequency change, and the array has infinite bandwidth. When the angle of incidence of the incoming wavefront changes, the number of delays at each element must be changed accordingly, in order to maintain the bandwidth. The previous discussion is valid for either a CW or a pulsed incidence signal. For the pulsed incidence signal, the time-delay feed preserves the shape of the pulse without any distortion.

When constant phase-type phasers (phasers whose phase shift is independent of frequency) are used at each element, the output phase distribution of the feed matches that of the incident phase front *only* at *one* frequency f_0 and for a particular incidence angle θ_0 . At a different frequency f_1 , the output phase distribution of the feed network remains fixed; hence the array is phased to receive at a different incidence angle θ_1 . The amount of beam squint with frequency is given by the following relationship:

$$f_1 \sin\theta_1 = f_0 \sin\theta_0 \quad (20-52)$$

For a small change in frequency, Eq. 20-52 shows that the change in scan angle is given by

$$\Delta\theta_0 = -\left(\frac{\Delta f}{f}\right)\tan\theta_0 \quad (20-53)$$

The previous expression shows that the amount of beam squint depends upon the original scan angle as well as on the percent frequency change. At broadside ($\theta_0 = 0$), there is no scanning, regardless of the amount of change in frequency, and the array has infinite bandwidth. When the beam scans away from broadside, the amount of beam squint with frequency increases with scan angle. Therefore, the bandwidth of an array must be specified in terms of the desired maximum scan angle. For most practical applications, the desired maximum scan angle is $\pm 60^\circ$ from array broadside. If we assume that the maximum allowable beam squint is ± 1 quarter beamwidth away from the desired direction, corresponding to a one-way gain loss of 0.7 dB, the bandwidth of the array is given by

$$\text{Bandwidth (percent)} = \text{beamwidth}(\circ) \quad (20-54)$$

Series Feed Effects

When a feed other than an equal length parallel feed is used, phase errors due to the feed alone are produced across the array aperture. An example is the series feed shown in Figure 20-18*b*. The total phase shift across the length of the feed is $\Phi = (2\pi/\lambda)L$ rad with free space propagation assumed. When the frequency is changed, the change in phase across the array aperture will be

$$\Delta\Phi = \frac{2\pi L}{c}\Delta f \quad (20-55)$$

This linear change in phase across the aperture scans the beam just as phasers would. To observe just how far the beam is scanned, examine the way in which an aperture is scanned with phase. For a given scan angle θ_0 the required phase across the array is

$$\psi = \frac{2\pi L}{\lambda}\sin\theta_0 \quad (20-56)$$

The required change in ψ for a change in scan angle is

$$\frac{d\psi}{d\theta} = \frac{2\pi L}{\lambda}\cos\theta \quad (20-57)$$

or

$$\Delta\psi = \frac{2\pi L}{\lambda}\cos\theta_0\Delta\theta_0 = \frac{2\pi Lf}{c}\cos\theta_0\Delta\theta_0 \quad (20-58)$$

when the change in phase across the array is induced by the feed, $\Delta\psi = \Delta\Phi$, or

$$\frac{2\pi L}{c}\Delta f = \frac{2\pi Lf}{c}\cos\theta_0\Delta\theta_0 \quad (20-59)$$

Hence the amount of beam scan for a change in frequency is given by

$$\Delta\theta_0 = \left(\frac{\Delta f}{f}\right)\frac{1}{\cos\theta_0} \quad (20-60)$$

For beams scanned in the direction of the feed load, the phase change between the elements due to the feed line is partially offset by the required phase change needed to maintain the beam position at a given scan angle θ_0 . In fact, for an endfire beam toward the load, the required phase change with frequency needed to maintain the beam at end free is exactly the same as that due to the feed line. Hence, the bandwidth of a series feed at the end-free beam position toward the load is infinite. However, this endfire beam position is not practical to implement because of the drastic reduction in the gain of the radiating elements. Furthermore, in most applications the array is required to scan in both directions from broadside. For beams scanned in the direction of the feed input, the phase changes in space and the feed line become additive, instead of canceling each other. As a result, the beam scans rapidly with frequency. For a 60° scan angle, the scanning caused by the feed alone is slightly greater than that caused by the aperture. Therefore, the bandwidth of the series feed is essentially half of that of the equal line length parallel feed.

In waveguide, the signal propagates more slowly; this is equivalent to having a longer feed line. This causes the feed to scan more rapidly, and the series feed performance degrades accordingly. The slowing in waveguide is proportional to λ/λ_g , and the array scans by

$$\Delta\theta \equiv \frac{\lambda_g}{\lambda} \left(\frac{\Delta f}{f} \right) \frac{1}{\cos\theta_0} \text{ rad} \quad (20-61)$$

For $\pm 60^\circ$ scanning from broadside, the bandwidth of a series feed under the same conditions as that of a parallel feed is given by

$$\text{Bandwidth (percent)} = \frac{1}{\left(1 + \frac{\lambda_g}{\lambda} \right)} \text{ beamwidth } (^\circ) \quad (20-62)$$

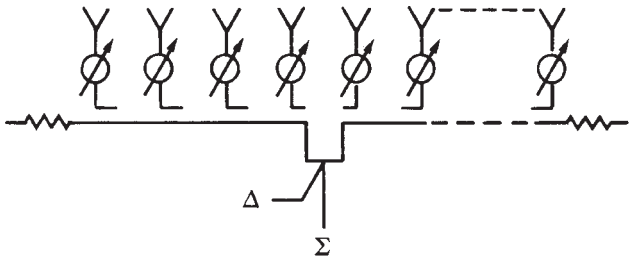
A center-fed series array,³⁶ as shown in Figure 20-26a, may be thought of as two end-fed series arrays. At band center the phase shifters are set so that all the elements radiate in phase. As the frequency is changed, each half scans in the opposite direction because of the gabled phase slope, as shown in Figure 20-26b. For the broadside beam case, this results in a broadening of the beam with no change in direction. If the two beams were to move far enough apart (Figure 20-27), the beams would split. However, by restricting the bandwidth to reasonable values, the result is a loss in gain caused by the beam broadening.

When the array is scanned away from broadside, aperture scanning due to the phaser (constant phase with frequency) is also introduced, and it is superimposed on the feed scanning. As shown in Figure 20-27 for one-half of the array (solid curve) aperture scanning and feed scanning tend to cancel each other, while for the other half (dashed curve) they tend to reinforce each other. The performance of the center-fed array is worse than that of the parallel-fed array at broadside but quite comparable at large scan angles. At a scan angle of 60° , the center-fed array will have approximately $\frac{1}{4}$ dB more loss than the parallel-fed array. For simplicity, the same bandwidth formula Eq. 20-54 can be used for both feeds.

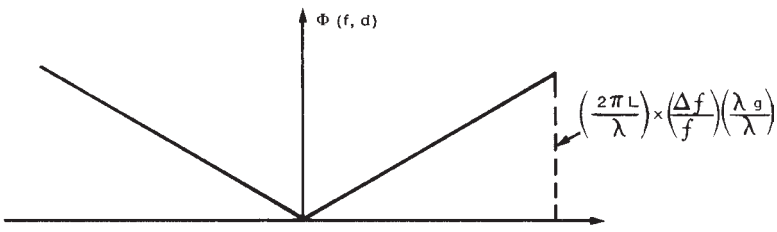
The previous discussion has assumed that the center feed used nondispersive transmission with free-space propagation. If waveguide is used, the performance of the center-fed array degrades and the bandwidth is given by

$$\text{Bandwidth (percent)} = \left(\frac{\lambda}{\lambda_g} \right) \text{ beamwidth } (^\circ) \quad (20-63)$$

Amplitude and phase errors also affect sidelobes.^{37,38}

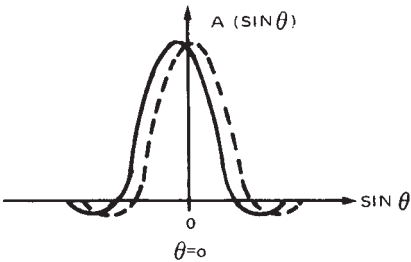


(a) CENTER-FED SERIES FEED

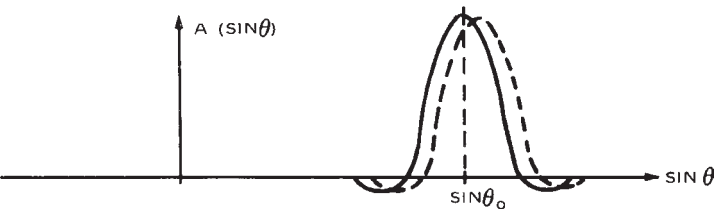


(b) GABLED PHASE DUE TO FEED

FIGURE 20-26 Center-fed series array



(a) FEED EFFECTS AT BROADSIDE



(b) APERTURE AND FEED EFFECTS

FIGURE 20-27 Aperture scanning superimposed on feed scanning

REFERENCES

1. S. Silver, *Microwave Antenna Theory and Design* (New York: McGraw-Hill, 1949).
2. S. A. Schelkunoff, "A Mathematical Theory of Linear Arrays," *Bell Syst. Tech. J.*, vol. 22 (1943).
3. R. C. Hansen, *Phased Array Antennas* (New York: John Wiley & Sons, 1998).
4. M. S. Smith and Y. C. Guo, "A Comparison of Methods for Randomizing Phase Quantization Errors in Phased Arrays," *IEEE Trans. Antennas and Propagat.*, vol. AP-31 (November 1983): 821–827.
5. R. C. Hansen and G. G. Charlton, "Subarray Quantization Lobe Decollimation," *IEEE Trans. Antennas and Propagat.*, vol. AP-47 (August 1999): 1237–1239.
6. E. D. Sharp, "Triangular Arrangement of Planar Array Elements That Reduces Number Needed," *IRE Trans. Antennas Propagat.*, vol. AP-9 (March 1961): 126–129.
7. T. T. Taylor, "Design of Line-Source Antennas for Narrow Beamwidth and Low Sidelobes," *IRE Trans. Antennas Propagat.*, vol. AP-3 (1955): 16–28.
8. C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beamwidth and Sidelobe Level," *IRE Proc.*, vol. 34 (1946): 335–348.
9. E. T. Bayliss, "Design of Monopulse Antenna Difference Patterns with Low Sidelobes," *Bell Syst. Tech. J.*, (May–June 1968): 623–650.
10. T. T. Taylor, "Design of Circular Apertures for Narrow Beamwidths and Low Sidelobes," *IRE Trans. Antennas Propagat.*, vol. AP-8 (January 1960): 17–22.
11. R. C. Hansen, "Tables of Taylor Distributions for Circular Aperture Antennas," *IRE Trans. Antennas Propagat.*, vol. AP-8 (January 1960): 23–27.
12. R. C. Hansen, "A One-Parameter Circular Aperture Distribution with Narrow Beamwidth and Low Sidelobes," *IEEE Trans. Antennas and Propagat.*, vol. AP-24 (July 1976): 477–480.
13. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design—Second Edition* (New York: John Wiley & Sons, 1998).
14. E. A. Wolff, *Antenna Analysis* (New York: John Wiley & Sons, 1966).
15. N. Amitay, V. Galindo, and C. P. Wu, *Theory and Analysis of Phased Array Antennas* (New York: Wiley-Interscience, 1972).
16. R. Janaswamy and D. H. Schaubert, "Analysis of the Tapered Slot Antenna," *IEEE Trans. Antennas and Propagat.*, vol. AP-35 (September 1987): 1058–1065.
17. D. H. Schaubert, "Wide-Bandwidth Radiation from Arrays of Endfire Tapered Slot Antennas," *Ultra-Wideband, Short-Pulse Electromagnetics*, H. L. Bertoni, L. Carin, and L. B. Felsen, (eds.), Plenum (1993): 157–165.
18. H. F. Lee and W. Chen, (eds.), *Advances in Microstrip and Printed Antennas* (New York: John Wiley & Sons, 1997).
19. R. C. Hansen, "A Critical Examination of Receive and Transmit Scan Element Pattern for Phased Arrays," *ACES Journal*, vol. 21 (November 2006): 299–304.
20. R. C. Van Wagoner and R. C. Hansen, "Measurement of Phased Array Scan Impedance by Load Pull," *Electronics Ltrs.*, vol. 39 (July 24, 2003): 1101.
21. P. W. Hannan and M. A. Balfour, "Simulation of a Phased Array Antenna in a Waveguide," *IEEE Trans. Antennas Propagat.*, vol. AP-13, no. 3 (May 1965): 342–353.
22. H. A. Wheeler, "A Survey of the Simulator Technique for Designing a Radiating Element," *Phased Array Antennas*, Oliner and Knittel, (eds.) (1972): 132–148.
23. R. C. Hansen, "A Gibbsian Model for Finite Scanned Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-44 (February 1996): 243–247.
24. R. C. Hansen and D. Gammon, "Standing Waves in Scan Impedance of Finite Scanned Arrays," *Microwave and Optical Tech. Ltrs.*, vol. 8 (March 1995): 175–179.
25. R. C. Hansen and D. Gammon, "Standing Waves in Scan Impedance: E-Plane Finite Array," *Microwave and Optical Tech. Ltrs.*, vol. 11, (January 1996) 26–32.

26. A. R. Lopez, "Monopulse Networks for Series Feeding an Array Antenna," *IEEE Antennas Propagat. Int. Symp. Dig.* (1967).
27. R. C. Hansen, "Phase and Delay in Corporate-Fed Arrays," *IEEE AP Magazine*, vol. 44 (April 2002): 24–29.
28. J. Butler and R. Lowe, "Beamforming Matrix Simplifies Design of Electronically Scanned Antennas," *Electron. Design*, vol. 9, no. 7 (April 1961): 170–173.
29. W. Rotman, "Wide Angle Scanning with Microwave Double-Layer Pillboxes," *IRE Trans. Antennas Propagat.*, vol. AP-6 (1958): 96–106.
30. R. C. Hansen, "Design Trades for Rotman Lenses," *IEEE Trans. Antennas Propagat.*, vol. AP-39 (April 1991): 464–472.
31. P. Barton, "Digital Beam Forming for Radar," *Proc. IEE*, vol. 127, Pt. F (August 1980): 266–277.
32. R. G. Pridham and R. A. Mucci, "Digital Interpolation Beamforming for Low-Pass and Bandpass Signals," *Proc. IEEE*, vol. 67 (June 1979): 904–919.
33. D. E. Dudgeon, "Fundamentals of Digital Array Processing," *Proc. IEEE*, vol. 65 (June 1977): 898–904.
34. J. Frank, "Bandwidth Criteria for Phased Array Antennas," *Proc. Phased Array Antenna Symp.* (June 1970).
35. R. Tang, "Survey of Time-Delay Beam Steering Techniques," *Proc. Phased Array Antenna Symp.*, Polytechnic Institute of Brooklyn, Brooklyn, NY, 1970.
36. R. R. Kinsey and A. L. Horvath, "Transient Response of Center-Fed Series Feed Array," *Proc. Phased Array Antenna Symp.*, Polytechnic Institute of Brooklyn, Brooklyn, NY, June 1970.
37. L. A. Rondonelli, "Effects of Random Errors on the Performance of Antenna Arrays of Many Elements," *IRE National Convention Record* (1959): 174–189.
38. K. R. Carver, W. K. Cooper, and W. L. Stutzman, "Beam-Pointing Errors of Planar-Phased Arrays," *IEEE Trans. Antennas Propagat.*, AP-21 (March 1973): 199–202.

Chapter 21

Array Phase Shifters: Theory and Technology

Robert R. Romanofsky
NASA Glenn Research Center

CONTENTS

21.1	INTRODUCTION	21-2
21.2	SEMICONDUCTOR.	21-4
21.3	THIN FILM FERROELECTRIC	21-10
21.4	MEMS	21-13
21.5	SLOW-WAVE.	21-14
21.6	FERRITE.	21-16
21.7	SUPERCONDUCTING APPLICATIONS	21-18
21.8	EFFECT OF PHASE SHIFTER BEHAVIOR ON PHASED ARRAY BIT ERROR RATE.	21-19
21.9	SUMMARY	21-22

21.1 INTRODUCTION

While there are a myriad of applications for microwave phase shifters in instrumentation and metrology, power combining, amplifier linearization, and so on, the most prevalent use is in scanning phased-array antennas. And while this market continues to be dominated by military radar and tracking platforms, many commercial applications have emerged in the past decade or so. These new and potential applications span low Earth-orbiting communications satellite constellations and collision warning radar, an aspect of the Intelligent Vehicle Highway System or Automated Highway System. In any case, the phase shifters represent a considerable portion of the overall antenna cost, with some estimates approaching 40 percent for receive arrays. Ferrite phase shifters continue to be the workhorse in military-phased arrays, and while there have been advances in thin film ferrite devices, the review of this device technology in the previous edition of this book is still highly relevant. This chapter will focus on three types of phase shifters that have matured in the past decade: GaAs MESFET monolithic microwave integrated circuit (MMIC), micro-electromechanical systems (MEMS), and thin film ferroelectric-based devices. A brief review of some novel devices including thin film ferrite phase shifters and superconducting switches for phase shifter applications will be provided. Finally, the effects of modulo 2π phase shift limitations, phase errors, and transient response on bit error rate degradation will be considered.

Applications

Military applications emphasize ground-based systems for early warning radar, missile defense, and space surveillance. Most of the systems employ ferrite phase-shifter technology, but several examples of GaAs MMIC-based arrays exist. Space applications, including civilian space applications, include synthetic aperture radar and satellite communications. Another burgeoning commercial application is collision warning and collision avoidance radar for the Intelligent Vehicle highway System.

Military Phased Array Antennas Table 21-1 summarizes some of the military phased array radar systems.^{1,2} Note that many of the arrays developed during the last three decades had production runs of over 50 and that even one radar system can necessitate a very large number of phase shifters.

The last two systems use MMIC modules. The Theater High Altitude Area Defense (THAAD) ground-based radar required over 60,000 MMIC phase shifter chips. The program demonstrated a per-module cost of about \$1,000. The Counter Battery Radar (COBRA) artillery and mortar weapon-locating system required more than 8000 MMIC modules.

Commercial Satellite Communications Customers for commercial satellite services include remote or mobile data-intensive professionals requiring fast downloading and uploading capabilities. Example users are banks, businesses using video-conferencing, oil drilling platforms, medical evacuation helicopters, airliners, cruise ships and the like. In developing countries, where a communications infrastructure is close to nonexistent, corporations are in need of ways and means to communicate. However, the greatest growth in demand may be fueled by consumer (residential) requirements such as: distance learning, electronic mail, home shopping, telecommuting, and entertainment services such as high definition television and video phones. Compared to geostationary (GEO) satellites, LEO satellites offer at least three major advantages. First, they orbit at an altitude generally below 1000 km instead of about 35,000 km. Since signal loss is proportional to distance squared,

TABLE 21-1 Phased Array System Examples (Reprinted with permission from *Microwave Journal*, 40, #5, (May 1997): 288–294.)

System	Frequency Band	Number Manufactured	Phase Shifters per Array	Elements Manufactured	Manufacturer
AN/TPN-25	X	18	824	14,850	Raytheon
AN/GPN-22	X	60	443	26,580	Raytheon
Cobra Dane	L	1	34,769	34,769	Raytheon
Pave Paws	UHF	4	2,667	21,416	Raytheon
BMEWS	UHF	2	3,584	17,920	Raytheon
Cobra Judy	–	1	12,288	12,288	Raytheon
Patriot	C	173	5000	865,000	Raytheon
AEGIS/SPY-1	S	234	4000	936,000	Raytheon
B-1	X	100	1,526	152,600	Northrup
AN/TPQ-37	S	102	359	36,618	Hughes
Flap Lid	X	>100	10,000	>1 million	USSR
THAAD	X	2.5	25,344	63,360	Raytheon
COBRA	C	3	2,700	8,100	Lockheed

an automatic power savings of about 30 dB occurs. This permits substantially smaller Earth terminals. Second, their proximity provides a nearly imperceptible propagation delay just like terrestrial systems, instead of the 0.25 second delay associated with GEO satellites. While this may be nothing more than a nuisance for voice services, it causes technical problems with higher data rates (for example, computer networking) and handshaking (for example, ATM switching). Third, there is potential for significantly reduced launch costs. For practical, aesthetic, or technical (agility and reliability) reasons, scanning phased arrays seems to be the lynchpin. System architectures tend to place considerable burden on the space segment, allowing the use of relatively small (for example, < 1 m) Earth terminals, to support data rates from perhaps 2.048 to 155.53 MBPS. These quixotic visions have yet to be realized, but could become tractable given a *low-cost* phased array solution.

There are also high data rate geostationary (GEO) platforms such as Spaceway (Hughes). Mobile platforms requesting service need some type of articulated antenna to track the satellite, both because of variable latitude and longitude, and to compensate for pitch, roll, and yaw. Again, the low-cost phased array seems an elusive solution.

Automated Highway Systems At the turn of the millennium, there were about 140 million automobiles just in the United States. As populations grow, highway traffic expands, but construction costs and available real-estate prevent the highway system from keeping pace. The Intelligent Vehicle Highway System, especially in the context of intelligent cruise control, collision avoidance radar, and electronic tolling is a solution to optimize traffic flow and reduce flawed decision making. For a collision warning application, the phased array can be relatively small. Preliminary specifications suggest an operating frequency of 77 GHz with a 1.5° by 6° beamwidth, a 10 Hz scan rate, and 10 mW output power.³ Electronically-steered arrays enhance collision avoidance radar both because of the beam-pointing precision requirements and the need to essentially see around corners.

The specified transmitter power corresponds to a timely warning for detecting a human, with a 1 m^2 radar cross-section, at a 300 m range.⁴

Phase Shifter Characteristics Evolving high data rate communications systems demand greater attention to subtle aspects of information theory and electromagnetic engineering. As the ratio of signaling bandwidth to carrier frequency decreases, less familiar phenomenon can influence system performance. And, new coding techniques are pushing channel capacity ever closer to the Shannon limit.⁵ Some interesting effects are expected to appear if the trend toward wide-band scanning phased array antennas and efficient high-speed modulators continues.⁶ For example, in a phased array antenna inter-element spacing, the physical size of the array, and the steering vector can conspire to introduce pulse distortion from group delay, inter-symbol interference, and beam squinting.^{7,8} And the operating point of the amplifiers can affect the bit error rate depending on the modulation type and the number of carriers. Naturally one wants the phased array to operate as efficiently as possible given power limitations and thermal management problems. This desire necessitates that the power amplifiers operate in a nonlinear region near saturation. Nonlinear effects cause amplitude-to-amplitude modulation (AM/AM) and amplitude-to-phase modulation (AM/PM) distortion. The net effect of AM/AM distortion is to alternately compress and expand the signal constellation. The net effect of AM/PM conversion is a rotation of the signal constellation.⁷ In a receive array, the third order intercept of the low noise amplifiers largely determines inter-modulation distortion and heat dissipation.⁹ But the phase shifter insertion loss envelope and phase accuracy are also key factors influencing array performance. Phase shifters typically follow low noise amplifiers in a receive array and precede power amplifiers in a transmit array. Since the phase shifter's insertion loss depends on its phase setting and since its switching action represents some finite time domain response, its potential contribution to bit error rate degradation cannot generally be ignored. There will always be some effects in any phase-shift keyed (PSK) modulation system, the degree to which depends on the steering vector update rate and data rate. A long switching time also increases minimum radar range. Besides these issues, the satellite communication market's desire to install tracking terminals on commercial mobile platforms, even at small office/home office and residential sites, has inspired the search for inexpensive phase shifters and affordable phased arrays. In practice, system constraints on chip size, power handling, drive power, insertion loss, bandwidth, phase error, transient response, and cost dictate particular device designs.

21.2 SEMICONDUCTOR

Semiconductor phase shifters, based primarily on GaAs, but also on SiGe and InP, have enjoyed steady progress for the past two decades. Their small size and relatively low power consumption compared to ferrite devices has created new insertion opportunities. Many possible circuit topologies, using diode or FET switches in various configurations, exist.

High-Pass/Low-Pass

In principle, any variable reactance in series or shunt across a transmission line can be used to introduce phase shift. A high-pass/low-pass phase shifter π network using discrete capacitors and inductors is shown in Figure 21-1. In the high-pass configuration shown, a relative delay is realized. In the opposite configuration, with all SPDT switches toggled, the low-pass circuit represents a relative phase advance.¹⁰ It can be shown that if the circuit is matched, $X = 2B/(1 + B^2)$ and the insertion phase is $\tan^{-1}(2B/(B^2 - 1))$.¹¹ The switches can be implemented with PIN diodes or MESFETs, which will be discussed later on in this section. It is possible to realize a phase shift of 180° with about 20% bandwidth.

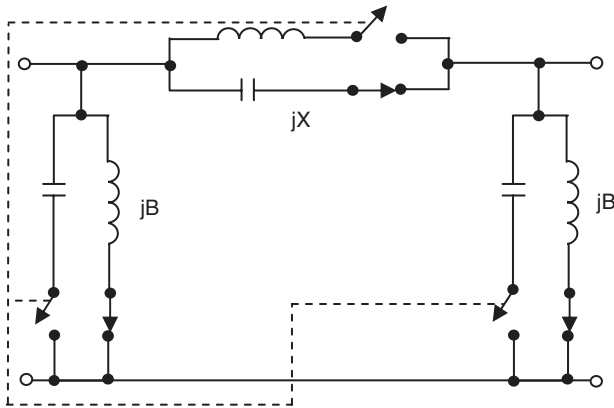


FIGURE 21-1 Topology for a high-pass/low-pass π network phase shifter. A dual “T” network realization is also possible.

Loaded Line

Another type of phase shifter generally used for achieving 22.5 to 45 degrees of phase shift is the loaded-line.¹² A schematic of this is shown in Figure 21-2. Ideally, reactive loads spaced one-quarter wavelength apart across a transmission line to effect phase shift. The purpose of the second shunt susceptance (jB) is to cause a reflection, which will at least partially cancel the reflection from the first shunt susceptance (jB).

By equating the ABCD matrix of Figure 21-2 to an equivalent section of transmission line with electrical length θ_L radians and characteristic impedance Z_0 Ohms as given in Eq. 1,

$$\begin{pmatrix} 1 & 0 \\ jB & 1 \end{pmatrix} \begin{pmatrix} 0 & jZ_0 \\ j/Z_0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ jB & 1 \end{pmatrix} = \begin{pmatrix} \cos(\theta_L) & jZ \sin(\theta_L) \\ j\sin(\theta_L)/Z & \cos(\theta_L) \end{pmatrix} \quad (21-1)$$

we obtain

$$Z = Z_0/[1 - (BZ_0)^2]^{1/2} \quad (21-2)$$

and

$$\theta_L = \cos^{-1}(-BZ_0) \quad (21-3)$$

If the susceptance is capacitive, the phase velocity is decreased; if the susceptance is inductive, the phase velocity is increased. Loaded line phase shifters are inherently

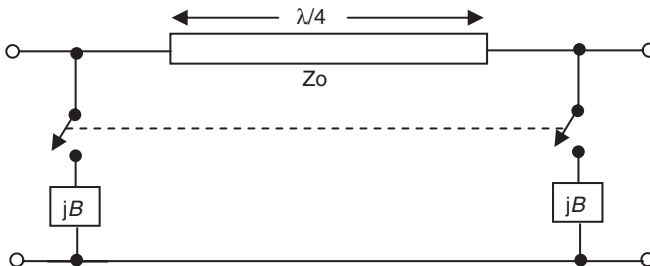


FIGURE 21-2 Loaded-line phase shifter

narrow-band and produce a constant phase shift versus frequency response. The phase versus frequency response is generally not as flat as the high-pass/low-pass type.

Switched Line

The switched-line phase shifter is yet another popular type, and is intuitively easy to understand. SPDT switches are used to toggle between transmission lines with different path lengths. As opposed to the types discussed previously, this one is a true time-delay device. That is, it provides a phase response (φ) proportional to frequency (ω). Since time delay, $\tau = -d\varphi/d\omega$ and φ is proportional to ω , τ is a constant over the bandwidth. A schematic is shown in Figure 21-3.

The differential phase shift is

$$\Delta\varphi \approx \beta(L2 - L1) \quad (21-4)$$

where β is the propagation constant of the transmission line. As usual, β equals the radian frequency ω divided by v_p , where v_p is the phase velocity. A point of caution needs to be made with regard to this design, however. Utilizing series diode switches, it is possible that the off path length and switch capacitance can conspire to create a through path in parallel with the on path, resulting in high insertion loss and abrupt phase change in band. Utilizing MESFETs in place of the SPDT switches shown, the on path could also experience high loss under certain conditions. If the MESFET is treated as a very small resistance in series with a pinch-off capacitance of 0.1 pF, the insertion loss (and phase) of the “on” path, say $L2$, will vary as shown in Figure 21-4 with $L2$ as a parameter.

The higher loss with increasing $L2$ is predominantly due to interaction with the switches, not dissipative loss. So $L1$ and $L2$ must be chosen with deference to the switch characteristics.

A photograph of a 4-bit GaAs monolithic phase shifter is shown in Figure 21-5.¹³ The chip size is approximately $5.5 \times 2.5 \times 0.15 \text{ mm}^3$. The 180° , 90° , and 45° bits are implemented using the switched line approach, whereas the 22.5° bit is realized with a loaded line. The operating frequency was 30 GHz with about a 10 percent bandwidth. Average insertion loss was $<10 \text{ dB}$ with an insertion loss envelope of about 2 dB.

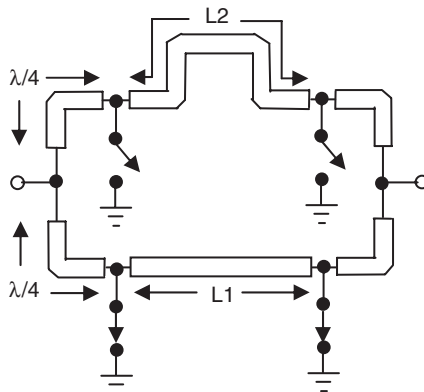


FIGURE 21-3 Switched-line phase shifter. Signal path is through the delay transmission line with the switches set as shown.

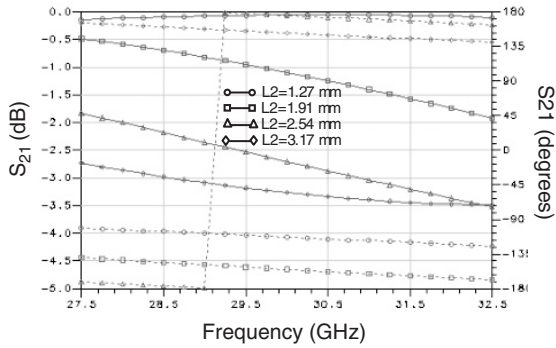


FIGURE 21-4 Modeled insertion loss and phase with path L2 as a parameter. The solid line is loss, and the dashed line is phase.

Beam Squint A clear advantage of a constant time-delay is that beam steering is independent of frequency. For simplicity, consider a linear array of N elements separated by distance d .¹⁴ The incremental phase shift between adjacent elements to form the beam at an angle θ from boresight is

$$\Delta\phi = (-2\pi d/c) f \sin(\theta) \quad (21-5)$$

where f is frequency and c is the speed of light in vacuum. If the frequency changes by Δf , the beam squint angle $\Delta\theta$ and phase deviation $\delta\phi$ are related by

$$(\Delta\phi + \delta\phi) = (-2\pi d/c) (f + \Delta f) \sin(\theta + \Delta\theta) \quad (21-6)$$

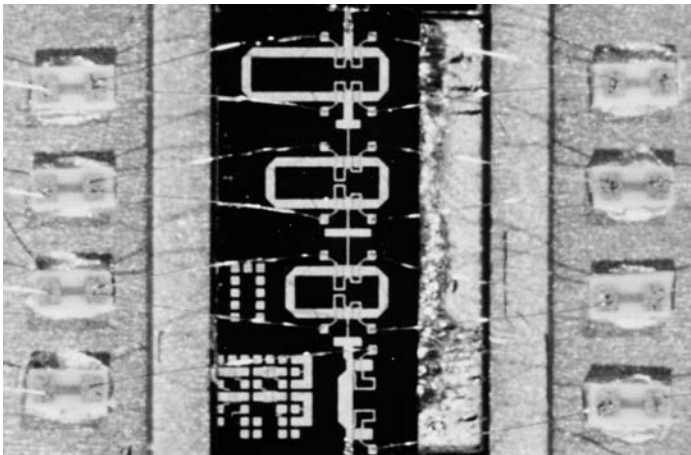


FIGURE 21-5 30 GHz GaAs MMIC phase shifter using $1 \times 400 \mu\text{m}$ switching FETs. The chip size is $5.5 \times 2.5 \times 0.15 \text{ mm}^3$ not including off-chip capacitors (after J. Geddes et al¹³ © IEEE 1987).

If the phase shifter insertion phase is proportional to frequency, the beam squint $\Delta\theta$ is zero. If the phase is independent of frequency, however, $\delta\phi$ is zero and equations (5) and (6) must be equal, yielding

$$f \sin(\theta) = (f + \Delta f) \sin(\theta + \Delta\theta) \quad (21-7)$$

which causes a beam squint angle

$$\Delta\theta = -\theta + \sin^{-1}[\sin(\theta)/(1 + \Delta f/f)] \quad (21-8)$$

Hence, wide scan angles and wide bandwidths correspond to significant scanning errors. In practice, phase shifters generally provide 0 to 2π phase shift. In electrically large arrays, the integer portion N of total delay required for true-time delay beam steering may be omitted, resulting in degraded performance. In radar systems, where narrow pulses are required for high resolution, there is a correspondingly wide frequency spectrum. The effect of this modulo- 2π problem is considered in detail later in this chapter.

Digital Control The distribution of amplifier and phase shifter control signals in a large phased array is a complex problem. Thousands of MMIC modules must be interconnected into the beam forming manifold. Some type of digital interface circuitry must be used to address individual phase shifters and decode their control signals. The multiplexed data rate to accommodate fast scanning may approach a Gb/s.¹⁵ In one instance, a GaAs optoelectronic integrated circuit was used to detect and demultiplex an optical serial control signal into 16 parallel electrical signals.¹⁶ Operation of a Ka-band, 4-bit phase shifter at 30 MHz clock speeds was demonstrated. Average optical power was 250 μ W. In another instance, optically controlled switching and X-band phase shifting was demonstrated such that the optically controlled GaAs FET interacted directly with the microwave signal.¹⁷ The FET was illuminated by the focused output of a 5 mW, 670 nm, InGaAlP laser diode.

Switching Q Key requirements for microwave phase shifters include bidirectional (or reciprocal) functionality, low power dissipation, and low insertion loss. The insertion loss of the switched line phase shifter in the previous section depends on the SPDT switch loss and transmission line loss. As we have seen, an SPDT switch can be realized by several possible combinations of SPST switches. Basically, the switch is intended to minimize insertion loss in one state and maximize isolation in the other. There are two fundamental configurations for SPST switches: as a series impedance or a shunt admittance. These are illustrated in Figure 21-6.

Defining insertion loss as the ratio of available power delivered by the generator, to the load in the absence of the switch, to the actual power dissipated by the load in the presence

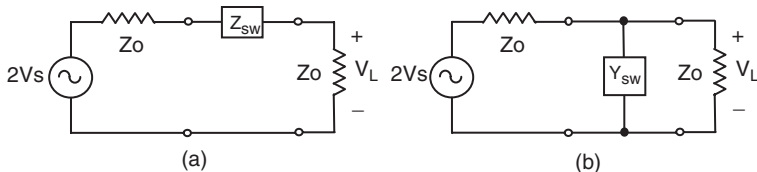


FIGURE 21-6 Equivalent circuit for series (a) and shunt (b) single pole, single through switch. The switch impedance and admittance are $Z_{sw} = R_{sw} + jX_{sw}$ and $Y_{sw} = G_{sw} + jB_{sw}$, respectively.

of the switch, the insertion loss for the series (L_{se}) and shunt (L_{sh}) switch is

$$L_{se} = 20 \log \left| 1 + \frac{R_{sw} + jX_{sw}}{2Z_0} \right| \text{ dB} \quad (21-9a)$$

$$L_{sh} = 20 \log \left| 1 + \frac{(G_{sw} + jB_{sw})Z_0}{2} \right| \text{ dB} \quad (21-9b)$$

In general, Z_{sw} and Y_{sw} , and Z_0 for that matter, are frequency dependent. The swing in Z_{sw} and Y_{sw} determine insertion loss and isolation. Of course, the diodes and MESFETs used to realize the switches are inherently nonlinear. Diodes are forward- and reverse-biased, in order to produce the maximum possible change in impedance. In the forward bias state, the impedance is small, but the diodes conduct substantial dc current. In the reverse bias state, a depletion region is formed. It's worth noting that according to well-accepted theories for junction capacitance C_j , as forward bias is increased, C_j grows unbounded. Of course, capacitance is a measure of stored charge, and as the depletion region shrinks towards zero dimension, the stored charge diminishes towards zero.¹⁸

GaAs MESFET devices have replaced diodes in many switching applications. Monolithic integration advantages and lower power consumption are incentives. A FET is basically a gate voltage controlled resistor. Insertion loss is largely dictated by channel resistance, and gate-to-source capacitance determines isolation. To increase isolation (high impedance off state), a short (inductive) section of a transmission line is connected between the source and drain to resonate out the pinch-off capacitance C_{off} .^{19,20} Ideal equivalent circuit models for on and off FETs are shown in Figure 21-7.

The resistor R_s is the total series resistance at pinch-off (undepleted channel resistance plus source and drain contact resistance). Sokolov²⁰ defined a figure of merit (Q) for switching FETs based on the work of Kurokawa and Schlosser²¹ for a device switching between two impedance states Z_1 and Z_2 , such that

$$Q^2 = \frac{|Z_1 - Z_2|^2}{r_1 r_2} = \frac{(R_{on} - R_s) + (\omega C)^{-2}}{R_{on} - R_s} \quad (21-10)$$

Here, r_1 and r_2 are the resistive components of the two impedance states. For the FET switch, $Z_1 = R_{on}$ and $Z_2 = R_s - j(\omega C_{off})^{-1}$. Assuming $(R_{on} - R_s) \ll (\omega C_{off})^{-1}$, a good approximation to Eq. 10 is

$$Q^2 = (R_{on} R_s)^{-1} (\omega C_{off})^{-2} = R_{off} / R_{on} \quad (21-11)$$

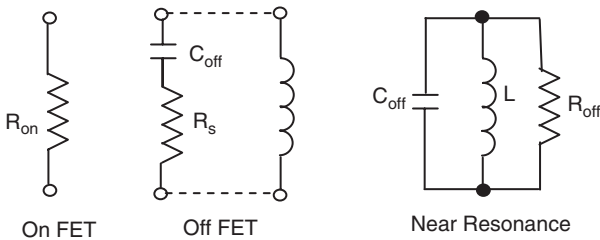


FIGURE 21-7 Ideal equivalent circuits for on and off FETs. The inductor L is a short section of a transmission line used to parallel resonate the RC combination in order to enhance the high impedance state.

Notice that Q degrades as the square of frequency. A high figure of merit requires small R_{on} , R_s , and C_{off} . Scaling the gate width allows a trade off between isolation and insertion loss. If the gate width is doubled, R_{on} and R_s are halved, C_{off} is doubled, and Q is invariant. Typical values of Q^2 at 10 and 30 GHz are about 1000 and 100, respectively. In normal operation, no bias is applied to the drain. A negative bias on the gate (relative to the source) pinches off the channel. Zero or positive gate bias turns the channel on. Additional information on switching Q and its relationship to loss is provided in the reference section.^{22,23}

21.3 THIN FILM FERROELECTRIC

Serious research on bulk (waveguide) ferroelectric phase shifters began in the 1960s but poorer performance in comparison to ferrite devices, coupled with very high tuning voltages, forestalled industrial acceptance. New thin film deposition techniques and novel circuit designs in the 1990s heralded a new generation of devices rivaling their semiconductor counterparts in essentially every figure-of-merit.

Materials and Basic Properties

In a ferroelectric material, such as $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$, applying a dc electric field shifts the position of the central Ti atom in the crystal, resulting in a net dipole moment on a macroscopic scale (ionic polarization). Analogous to ferromagnetic materials, domains form in the material where the dipoles are locally aligned (orientational polarization). The Ti ion can remain shifted even after the applied field is removed, and the domains can move in response to applied ac and dc fields. Microwave phase shifters are generally operated above the Curie temperature, in the paraelectric state. Ideally, this means operating without residual polarization and domain formation, thereby reducing hysteresis and dielectric loss. The Curie temperature, which represents a phase change to a nominally cubic crystal lattice, can be tailored for a specific operating temperature by adjusting the composition of $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$ (BST), where $0 < x < 1$, and for room temperature $x \approx 0.60$. Interest in ferroelectric-based agile microwave circuits is mounting because of their high power handling capability, negligible dc power consumption, and potential for low loss and cost. The ferroelectrics discussed here belong to the perovskite crystal family. The dielectric constant (ϵ_r) of single crystal SrTiO_3 , an incipient ferroelectric, can be depressed from about 20,000 to 2000 with a dc field of 10^4 V/cm at 4.4 K (breakdown voltage for the materials of interest here is $>10^5$ V/cm) and the loss tangent maintained below 0.1%. Thin films of SrTiO_3 exhibit $\tan\delta$ as poor as 0.01 with a peak relative dielectric constant of ≈ 5000 . The dielectric constant also tends to exhibit a broad maximum with temperature as opposed to bulk material. The differences in behavior have been attributed to residual domain wall motion, compositional inhomogeneities, interface layers between the film and electrodes, and lattice mismatch induced stress. Fundamental loss mechanisms have been considered in some situations.²⁴ Attempts to reduce $\tan\delta$ have included annealing, which tends to increase grain size, and the use of dopants, which may reduce free charge or otherwise affect loss mechanisms.^{25–27}

Several ferroelectric phase shifters have been developed with varying success. A stripline circuit with a BST capacitor provided a differential phase shift of 11° at X-band with a biasing field of 70 kV/cm.²⁸ In that same work, a center-wire bias waveguide phase shifter produced more than 360° of phase shift at Ku-band by changing the bias between the wire and waveguide walls from 0 to 2500 V. A planar microstrip phase shifter was reported that provided $20^\circ/\text{kV}$ at 2.65 GHz.²⁹ A phase shift of 165° at 2.4 GHz with only 3 dB loss and a bias of 250 V was obtained from a microstrip on a thin BST slab synthesized using a sol-gel technique.³⁰

There was also a 40 GHz phased array antenna that used radiating slots in waveguide and a BST film sintered onto a MgO substrate.³¹ Voltage applied across a periodic set of electrodes changed the dielectric constant of the BST from 700 to 1500 and a $\tan\delta$ of 0.05 was reported. A ferroelectric lens that uses BST slabs sandwiched between conducting plates was also proposed.³² The approaches advanced thus far have not been able to simultaneously address low cost, low loss, and small size. And in some cases, the impedance variation, due to widely changing permittivity, posed additional difficulties. Two promising implementations include coupled line and synthetic line-based devices, to be discussed next.

Coupled Microstripline

Another style of phase shifter uses coupled microstriplines as dc electrodes to polarize a thin ($\approx 0.4 \mu\text{m}$) ferroelectric film. With $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ electrodes and $2.0 \mu\text{m}$ thick SrTiO_3 films, this phase shifter produces a figure of merit approaching $120^\circ/\text{dB}$ at 40 K.³³ At room temperature, using Au electrodes and 400 nm thick $\text{Ba}_{1-x}\text{Sr}_x\text{TiO}_3$ films devices have demonstrated $\approx 70^\circ/\text{dB}$.³⁴ These planar phase shifters are fairly compact, low loss, easy to fabricate, and can provide 360° of phase shift with bias voltages under 350 V. Such devices can enhance conventional (direct radiating) phased array performance or enable a new type of reflectarray antenna.^{35,36} Only one control line is required per phase shifter, thereby simplifying array assembly. A variation is a hybrid X-band phase shifter consisting of four cascaded coupled microstrip lines, patterned over 400 nm thick laser-ablated $\text{Ba}_{0.50}\text{Sr}_{0.50}\text{TiO}_3$ films, followed by a switch.³⁷ The ferroelectric section provides (nominally) 180 degrees of analog phase shift. Basically, as a bias from 0 to 350 V is applied across the coupled line electrodes, the relative dielectric constant of the film tunes from about 2000 to 800, thereby modifying the propagation constant. The ferroelectric films are, of course, excellent dielectrics and the current draw is negligible so there is virtually no power consumption. Switching speed for these devices is limited by the external power supply. This is analogous to MESFET switches, wherein there is no drain bias. The variable resistance of a FET switch is a function of the population of majority carriers, as opposed to minority carriers in PIN diodes. (For example, in FET switches, speed is not limited by minority carrier lifetime, but rather the gate control circuitry). The beam lead GaAs diode switch is appended to the last coupled microstrip section and toggles between an open and virtual short circuit, realized with a quarter-wave radial stub. This results in a “digital” transition between a reflection coefficient with magnitude near unity and phase of ≈ 0 degrees and ≈ 180 degrees, respectively. Thus, a full 2π phase shift is possible. The average loss at X-band was 3.2 dB with a 10% bandwidth. The device is pictured in Figure 21-8. Note that the unmatched beam-lead diode contributed 0.5 to 0.75 dB to the overall insertion loss. In principle, this can be improved by “resonating-out” the diode capacitance.

Theory of Coupled-Line Type

A sketch of the cross-section of a coupled-line (multi-dielectric-layer) ferroelectric phase shifter is shown in Figure 21-9. Y_1 and Y_2 represent the admittance looking in the positive and negative y direction, respectively, from the charge plane. The thickness of the ferroelectric layer is h_1 while the host substrate has thickness h_2 .

By concentrating the fields in the odd mode, the phase shift per unit length is maximized, and by using the material in thin film form, the effects of high loss tangent are reduced.

The amount of phase shift can be increased by cascading coupled line sections. Though methods for calculating the propagation parameters of coupled transmission lines are well known, coupled lines on stratified substrates are difficult to analyze. And the high permittivity of the ferroelectric layer causes long computation time by full-wave electromagnetic

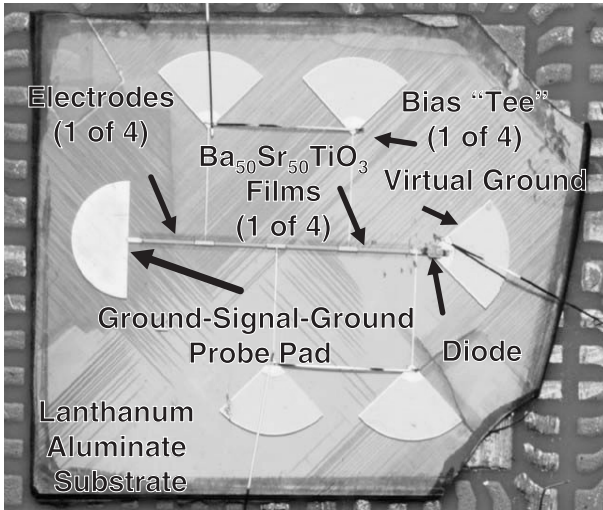


FIGURE 21-8 Hybrid X-band ferroelectric/semiconductor phase shifter on 0.5 mm thick lanthanum aluminate. The device measures $\approx 1 \times 1$ cm.

simulators because the geometry must be fractured into many thousands of cells. The multilayer structure has been analyzed using a computationally efficient variational method for calculating the complex propagation constant and characteristic impedance.³⁸ Space limitation prevents a full description of that work here. However, the method is quite general and can be used for multiple layers of various dielectrics or other types of transmission lines. For example, a multilayer microstrip can be analyzed by allowing the strip spacing (s) to become much greater than the effective substrate thickness (h) or strip width (w). Minimum loss occurs when current flows through one line and returns through the other (odd-mode),

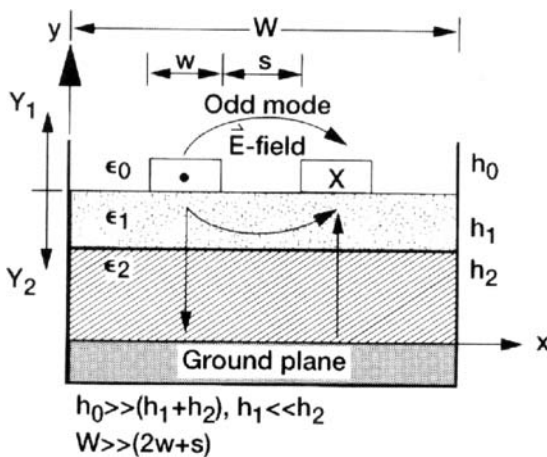


FIGURE 21-9 Cross-section of the coupled microstripline phase shifter, showing the odd-mode electric field configuration

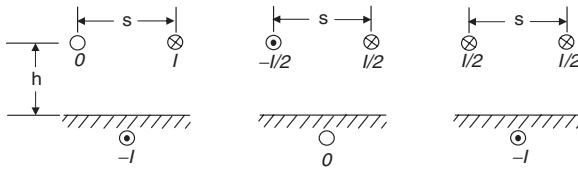


FIGURE 21-10 Coupled line modes: microstrip mode (left), $s \gg h, w$; odd-mode (middle); even-mode (right)

obviating the need for the ground plane. Maximum loss occurs when equal currents flow through the two lines and return through the ground plane (even-mode). Three basic modes are illustrated in Figure 21-10.

Note that the ferroelectric layer thickness is crucial to performance. In principle, the phase shift for a $2\text{ }\mu\text{m}$ thick film is 2.2 times greater than that of a $0.5\text{ }\mu\text{m}$ film. However maintaining the crystalline quality of the pulse laser ablated $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$ films past a thickness of $0.5\text{ }\mu\text{m}$ or so has proven to be difficult. Experiments have shown that the insertion phase shift is roughly proportional to $h_1^{0.67}$. In the case of cascaded coupled lines, increasing phase shift occurs at the expense of bandwidth since the structure resembles a multipole filter. Changing the dielectric constant of the ferroelectric film to change its insertion phase also modifies the pass band characteristic. Operation near band edges is to be avoided since that represents distorted transmission which degrades bit-error-rate performance. The device discussed in the previous section “Coupled Microstripline” helps alleviate this problem. The impedance matrix of the cascade network can be derived by traditional coupled line theory, using the superposition of even and odd mode excitation. Then an equivalent S-parameter model can be extracted and used to predict the pass-band characteristics of the phase shifter. A key advantage of this technology is the relatively large feature size. Active devices at the frequencies of interest here would necessitate submicron gate length GaAs FETs. The finest feature size associated with the coupled-line phase shifters is the electrode separation, typically $\approx 10\text{ }\mu\text{m}$. Whereas the GaAs FET performance is largely dictated by transconductance, and hence, carrier transit time across the gate region, the coupled-line phase shifters are static devices. The electrode gap separation determines the degree of electromagnetic coupling and the dc potential required to tune the film.

21.4 MEMS

Micro-electro-mechanical systems (MEMS) can essentially be used to replace PIN diode or FET switches in conventional phase shifters. A typical switch consists of a metal membrane bridge suspended several microns above a lower metal contact. Electrostatic attraction between the flexible membrane and lower contact is used to toggle the switch. A very thin dielectric coating (for example, Si_3N_4) may be used on the lower contact to eliminate a dc path and reduce “stiction,” or a tendency to latch (also known as Van der Waals forces). Larger bridge heights (gaps) reduce parasitic capacitance, but increase pull-down voltage and may contribute to more fatigue. Typical actuation voltages are between 20 and 100 V. An interesting advantage of MEMS switches is that they do not depend on the characteristics of the substrate. Hence, they may be fabricated on any material that is compatible with standard IC processing. A photograph of one type of MEMS switch is shown in Figure 21-11. Optimizing the switching Q amounts to minimizing the parasitic capacitance of the bridge in the off state (bridge up), and maximizing capacitance in the on state (bridge down). When the cantilever is down, the parallel plate capacitance between the metal membrane and the bottom electrode, C_{on} , is large and the switch behaves as a short circuit. Typical ratios of $C_{\text{on}}/C_{\text{off}}$ are ≈ 100 .

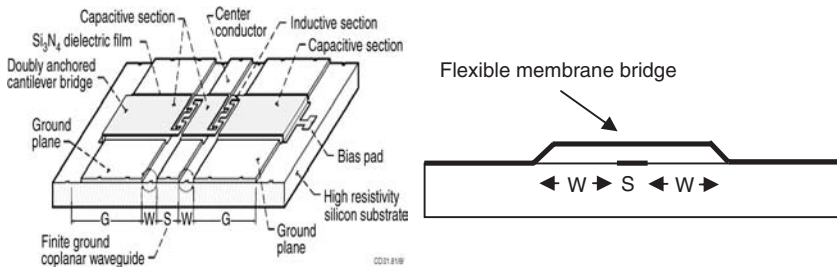


FIGURE 21-11 RF MEMS capacitive switch. MEMS devices incorporating capacitive\inductive sections allow the switch to be designed for minimum loss and maximum isolation over a wide frequency range (after M. Scardelletti⁴³ © IEEE 2002).

A number of distributed MEMS transmission line phase shifters have been developed.³⁹ An insertion loss of 4 dB at 60 GHz (for a full 360° phase shift) and 257°/dB at 50 GHz has been reported.⁴⁰⁻⁴¹ These excellent results have been obtained from MEMS bridge capacitors placed periodically along a coplanar waveguide (CPW) transmission line. The theory of such “slow-wave” circuits is discussed in the next section. Activating the MEMS increases the capacitive loading, thereby controlling the insertion phase. A mechanical analysis of operation was provided⁴² and the bridge spring constant was given as

$$k \approx 32Et^3w/(S + 2W)^3 + 8\sigma(1 - \nu)tw/(S + 2W)N/m \quad (21-12)$$

where E is Young’s modulus of the bridge material, t is the bridge thickness, w is the bridge width, S is the CPW center conductor width, W is the CPW center conductor-to-coplanar ground spacing, σ is the internal residual stress of the bridge, and ν is Poisson’s ratio. The pull-down voltage V_p is given as

$$V_p = [8\text{ kg}_0^3/(27\epsilon_0Sw)]^{1/2} V \quad (21-13)$$

where ϵ_0 is the free-space permittivity and g_0 is the zero-bias bridge height.

One issue with MEMS phase shifters is that packaging is perhaps more critical than with alternative technologies. Hydrocarbon contamination may compromise MEMS reliability so hermetic seals and careful processing are required. A summary of performance and implications for phased arrays are presented in the reference section.⁴⁴⁻⁴⁵

21.5 SLOW-WAVE

Periodically loaded transmission lines are used to realize band-pass, slow-wave circuits such that $v_p \ll c$. By using tunable loads or varactors, broad-band phase shifters can be designed.

One-Dimensional Periodic Structures

Consider a cascade of short sections of lossless transmission line shunted by a voltage variable capacitance C' , so as to form the infinite periodic structure illustrated in Figure 21-12. Each distributed transmission line section is modeled by an LC circuit and each unit cell consists of L , C , and C' . Let the separation of each unit cell be designated as s . Such periodic structures exhibit slow-wave behavior ($v_p \ll c$) and band-pass characteristics, like filters.

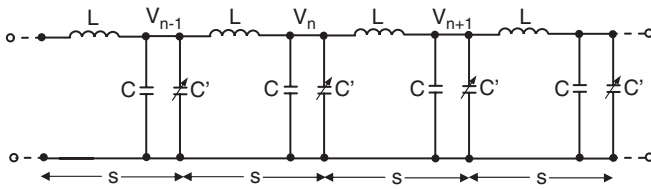


FIGURE 21-12 A chain of LCC' circuits representing an infinite, periodically loaded transmission line resulting in slow-wave propagation

The phenomenon is analogous to acoustic wave propagation in crystal structures. These types of circuits are often referred to as synthetic line, slow-wave, or distributed transmission line phase shifters.

A wave traveling through this periodic structure will only experience a phase shift from unit cell to unit cell, such that V_n is delayed relative to V_{n-1} as

$$V_n = V_{n-1} e^{-j\theta_s} \quad (21-14)$$

In general, θ_s could be complex to account for attenuation as well as phase shift. Summing currents leaving node n

$$0 = V_n(j\omega C + j\omega C') + (V_n - V_{n-1})(-j/(\omega L)) + (V_n - V_{n+1})(-j/(\omega L)) \quad (21-15)$$

Substituting Eq. 12 into Eq. 13 we arrive at

$$\cos(\theta_s) = \frac{1}{2}(2 - \omega^2 L(C + C')) \quad (21-16)$$

Requiring θ_s to be real to represent a propagating mode, letting $\cos(\theta_s) = \pm 1$, we find the structure of Figure 21-12 has a zero lower frequency cutoff ω_L (obvious by inspection) and an upper cutoff frequency ω_H corresponding to

$$\omega_H = 2/(L(C + C'))^{1/2} \quad (21-17)$$

Since $\theta_s = \beta s$, the insertion phase shift can be approximated knowing the extremes of the tuning range of C' and the number of cascaded sections. Slow wave phase shifters are, in principle, time delay devices (in other words, phase shift is proportional to frequency).

Ferroelectric Varactor

One type of slow wave phase shifter, based on ferroelectric “varactors” is shown in Figure 21-13. Coupled lines, as described in the previous section called “Theory of Coupled-Line Type,” are attached transversely to the propagation direction in a microstrip configuration.⁴⁶ A unit cell consists of two microstrip sections, with an intervening orthogonal coupled microstrip line terminated in a virtual open circuit. The open circuit is realized by cascading a nominally quarter-wave high-impedance microstrip line, having impedance Z_p , with a radial stub. The stub provides a very good approximation to a short circuit at its vertex, over almost an octave bandwidth, and the input impedance (Z_{in}) to the combination is approximately $35Z_p/\alpha$, where α (dB/ λ) is the attenuation of the Z_p line. Note that there is an optimal choice for Z_p , since as Z_p increases to maximize Z_{in} , α also increases due to conductor loss.

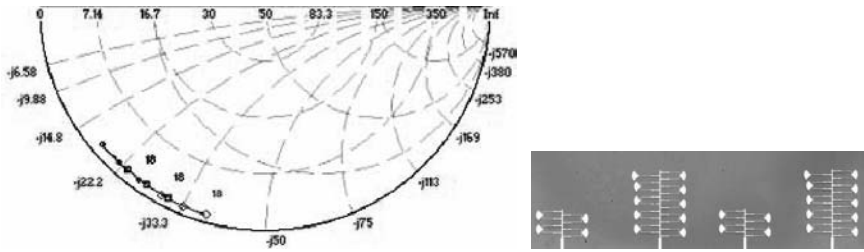


FIGURE 21-13 Modeled impedance variation of a 19-cell slow wave circuit from 18 to 20 GHz as the ferroelectric film dielectric constant varies from 2000 to 800 (left). The actual circuit (right) had a 0.6 mm cell size with 200 μm long coupled-line varactors.

For the reactances chosen and shown, the modeled dissipative loss was about 0.125 dB/unit cell. There is a trade-off between loss and phase shift. For a median varactor reactance of $-j500$, the calculated loss per unit cell is 0.003 dB, but the phase shift is also $\ll 1^\circ/\text{cell}$ and appears impractical. For a reactance of $-j5$, the insertion loss is about 8.6 dB/cell and is also impractical. For the K-band circuit shown, the varactors theoretically tuned from 0.45 pF (zero bias) to 0.25 pF (maximum bias), and 16 unit cells should have provided 360° of phase shift at 20 GHz. These results could be obtained using the techniques described in the subsection entitled “Coupled Microstripline.” The measured phase shift of the actual 19 cell structure was about 250° with an average loss of about 7.5 dB.

Excellent results have been obtained from slow wave circuits using parallel-plate ferroelectric varactors.^{47–49} The circuits operating principles are basically as described in Section 21.5. The device consists of a high impedance transmission line on sapphire, periodically loaded with $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$ capacitors spaced by distance s . Recently, $\approx 360^\circ$ phase shifters at K- and Ka-band exhibited an average loss of about 5 and 6 dB, respectively.⁵⁰ One advantage of the parallel plate approach is that conventional tuning voltages can be used (for example, ≈ 10 V as opposed to >100 V for coplanar structures). Another advantage is that circuits can be fabricated on convenient substrates like Si instead of exotic, high epsilon substrates like LaAlO_3 . Propagation in semiconductor substrates has also been discussed.⁵¹ The circuits are also extremely compact. A Ku-band version measuring 1×0.8 mm was reported.⁵⁰

Ferroelectric deposition methods (for example, pulse laser, sputtering, and so on) require an oxidizing atmosphere around 700°C to achieve high quality crystalline films. Hence, electrodes for parallel plate varactors, as opposed to the coplanar or interdigital structures described previously, must be impervious to high temperature oxidizing atmospheres. Low dielectric constant oxides at the electrode interface can drop most of the electric field and substantially reduce tuning. Pt is a popular electrode material choice.

21.6 FERRITE

The fundamental source of a material’s magnetic properties is the magnetic dipole moment associated with what is visualized as an electron spinning about its axis. In non-magnetic materials, the number of electrons with up spins and down spins is equal, so the net magnetic moment is negligible. In a magnetic material, one variety dominates and can be aligned by an external DC magnetic bias field H to generate a large magnetic moment, as long as thermal fluctuations aren’t too strong. The spin magnetic moment vector precesses about

the H -field vector at an angular frequency $\omega_m = 2\pi f_m$. Precession sense (i.e., polarization) depends on the direction of H . The precession frequency is proportional to H such that $f_m = \gamma H$, where $\gamma = 2.8 \text{ MHz/Oe}$ is the gyromagnetic constant. The propagation constant (β) in a ferromagnetic medium is

$$\beta \approx (2\pi\sqrt{\epsilon\mu'})/\lambda \quad (21-18)$$

where μ' is the real part of the permeability and ϵ is the dielectric constant of the ferrite. When the frequency f of a microwave magnetic field equals f_m , ferrimagnetic resonance occurs. This is the underlying cause for opposite circularly polarized waves experiencing different permeabilities, μ'_+ and μ'_- . In the absence of any loss mechanism, the material's susceptibility would become infinite at the gyromagnetic resonance frequency. The permeability can be tuned by changing M , which is the magnetization or magnetic moment. As the strength of the magnetic bias field increases, M will reach an upper limit called the saturation magnetization, denoted as M_s . Ferrite materials tend to have high loss below saturation at microwave frequencies. The choice of ferrite for a particular application is determined by selecting $\gamma 4\pi M_s$ sufficiently below the operating frequency f . When f_m is small compared to f , there is a significant difference between μ'_+ and μ'_- for the two polarization senses. (Recall that a linearly polarized wave can be disassociated into two counter-rotating circularly polarized components). With this condition, it can be shown that⁵²

$$\mu'_{\pm} = 1 \mp \gamma 4\pi M/f \quad (21-19)$$

and the corresponding phase shift is

$$\Delta\phi = 2\pi\sqrt{\epsilon} [(\mu'_+)^{1/2} - (\mu'_-)^{1/2}]L/\lambda \quad (21-20)$$

where L is the length of the ferrite section. In practice, the phase shift is implemented by reversing the direction of the magnetization vector.

As was alluded to in the introduction, ferrite phase shifters have tended to dominate the phased array landscape. The basic, nonreciprocal, latching phase shifter consists of a toroidal ferrite core in a section of waveguide. A drive wire passes through its center, which is used to magnetize the material. At this point it becomes latched at one of two remanent states, depending on the polarity of the bias. The amount of differential phase shift between the two states is determined by the toroid length. Another common device is the reciprocal, non-latching Reggia-Spencer phase shifter. A ferromagnetic bar is located axially in a section of waveguide, which is wound by a solenoid. The solenoid produces a longitudinal magnetic field that changes the permeability of the bar material. The reader is referred to the previous edition of this book and the references therein for complete descriptions. The performance of twin-toroid, dual-mode, and rotary-field ferrite phase shifters has been discussed elsewhere.⁵³

Microstrip Type

Our interest here will be a brief introduction to microstrip-type ferrite phase shifters. Planar microstrip geometries promise an economical and small-sized replacement for waveguide phase shifters. For example, a small analog X-band stripline on a ferrite slab yielded a figure-of-merit of 300°/dB with a 30 mW average drive power.⁵⁴ A renewed interest in this type of device was spurred on by the advancement of high-temperature superconductors, which are briefly discussed in the next section. The key advantage of superconducting transmission lines is the very low surface resistance compared to metallic conductors.

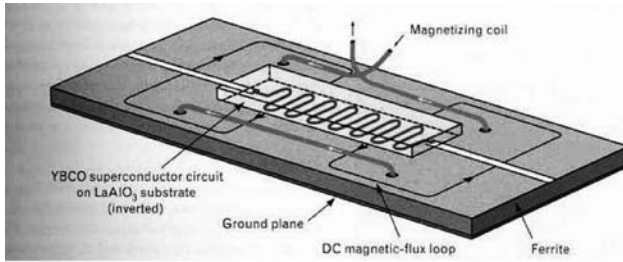


FIGURE 21-14 Geometry for a superconducting microstrip meander line phase shifter inverted and in close proximity to a tunable ferrite substrate. The device is about 1.3 cm long (after G. F. Dionne et al.⁵² © Lincoln Laboratory, MIT 1996).

Low-loss microstrip ferrite phase shifters were demonstrated by combining the superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ with polycrystalline yttrium iron garnet $\text{Y}_3\text{Fe}_5\text{O}_{12}$. The design confined the magnetic flux within the ferrite to avoid undesirable flux penetration into the superconductor, yet allowed gyromagnetic interaction with the microwave signal.⁵⁵⁻⁵⁶ A prototype phase shifter using a $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ microstrip meander line deposited on a LaAlO_3 substrate is shown in Figure 21-14.

A differential phase shift in excess of 200 degrees with an insertion loss generally below 2 dB from 6 to 12 GHz was reported.⁵² Besides the insertion loss advantage of the superconducting line, the $4\pi M_s$ values can increase by more than a factor of two when the material is cooled from 300 K to 77 K.

21.7 SUPERCONDUCTING APPLICATIONS

While conventional or low temperature superconductor based phase shifters have been developed to some extent,⁵⁷⁻⁵⁸ the discovery of so-called “high-temperature” ceramic superconductors in 1986⁵⁹ ushered in a new class of microwave components. High temperature superconducting (HTS) phase shifters are a subset of that class. The main attraction was dramatically reduced surface resistance compared to metallic conductors. The three primary HTS materials are $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, BiCaSrCuO_x , and $\text{Tl}_2\text{Ba}_2\text{Ca}_2\text{O}_x$ with respective transition temperatures (T_c) of 95 K, 110 K, and 122 K. Surface resistance (R_s) of these superconducting films at 77 K is on the order of $100 \mu\Omega$ at 10 GHz, compared to about $10 m\Omega$ for copper.

The surface impedance of a superconductor is derived using a two-fluid model, which maintains that the current is composed of both superconducting (paired) and normal electrons. The proportion varies such that at $T = 0$, all carriers are superconducting and at $T = T_c$, all carriers are normal. The two-fluid model predicts that the surface resistance of a superconductor is proportional to f^2 , as opposed to $f^{1/2}$ for a metallic conductor.⁶⁰ The strong dependence of R_s on frequency is a consequence of reactive voltage, associated with the oscillating supercurrent interacting with unpaired (normal) electrons to induce loss. The corresponding (inductive) surface reactance is $\omega\mu\lambda_L$, where λ_L is the London penetration depth. λ_L has a minimum value at $T = 0$ and diverges as T approaches T_c . The abrupt change in resistance and kinetic inductance ($\mu\lambda_L$) near T_c presents opportunities for novel detectors (for example, transition edge bolometers) and switching mechanisms.

Reflective switches can be based on driving a narrow superconducting bridge into the normal state by locally exceeding the superconductor’s critical temperature.⁶¹ Alternatively,

a magnetic field on the order of 100 G was used to cause a small bar of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ to transition from the superconducting to normal state.⁶² The effects of optical illumination have also been investigated.⁶²⁻⁶³

A novel phase shifter based on a HTS distributed Josephson inductance transmission line was reported.⁶⁴ Superconducting quantum interference devices (SQUIDS), consisting of a superconducting inductor shorted by a Josephson junction, were distributed along a length of superconducting transmission line. Each SQUID can be modeled as a variable inductor, which is tuned by applying a magnetic field. The circuit constitutes a dual of the structure discussed in Section 21.5. Slow-wave propagation in superconducting transmission lines has also been studied. For example, the dramatic increase in λ_L near T_c causes a large reduction in v_p if the film is very thin.⁶⁵

We conclude this section noting that the field of cryogenic electronics is growing and may be further stimulated by recent advances in compact, reliable, mechanical refrigerators. While a few esoteric systems have been created for space-based astrophysics, ultra-sensitive receivers, and medical diagnostics, there may also be niche applications in communications markets. Cryogenically cooled receivers for cellular phone towers are a good example. The enhanced sensitivity allows for increased cell-size. It is easy to speculate that applications for cryogenically cooled agile antennas, and hence phase shifters, may also eventually exist. Digital beam-forming systems based on Josephson technology are also being contemplated.

21.8 EFFECT OF PHASE SHIFTER BEHAVIOR ON PHASED ARRAY BIT ERROR RATE

Phase shifters' phase errors can cause some signal distortion, and phase transients can cause beam pattern degradation during direction switching. Signal models have been established for both static and transient cases. Moreover, modulo 2π effects cause intersymbol interference (ISI). This section is based on the work reported in this area.⁶⁶

Modulo 2π Effects

The composite far-field signal formed by a phased array is the summation of individual contributions from each elemental radiator. Radiation path lengths to the far-field observation point are different for each elemental radiator on the array surface. To form a cophasal beam, modulo- 2π compensation is provided by the phase shifters. Integer-multiples of 2π are generally not provided, and for an electrically large array, this leads to inter-symbol interference (ISI) in phase-shift keyed modulation formats. The ISI is formed because different delays from each signal component destructively interfere, as illustrated in Figure 21-15, for Binary Phase Shift Keyed waveforms. ISI results in composite (total) signal amplitude loss at the symbol boundary.

As an example, for the reflectarray³⁵⁻³⁶ operating at $f = 26.5$ GHz with a bit rate of 1.325 Gbps, the BER degradation due to ISI is proportional to the symbol rate and the loss ranges are from about 1 dB to around 2 dB in E_b/N_0 , depending on original E_b/N_0 , for BPSK, QPSK, 8PSK, and 16QAM.

Phase Errors

Each phase shifter assigned to each element is designed to shift the signal phase by an amount determined by the steering angles of the antenna. In practice, however, there are errors in the phase shifters. Assume that the phase errors are uniformly distributed in

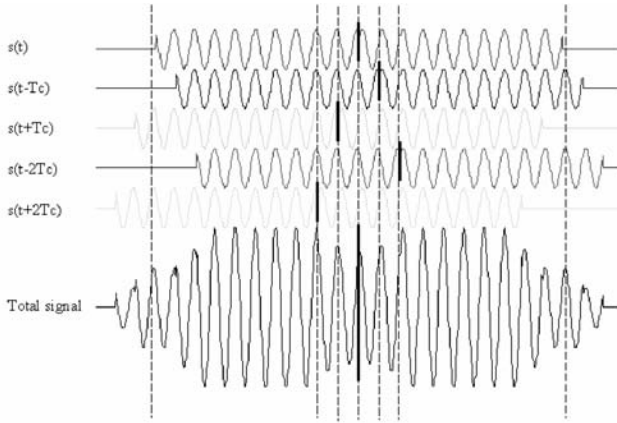


FIGURE 21-15 Formation of ISI due to different delays in signal components. The carrier period is designated T_c . Heavy vertical lines are symbol boundaries (after F. Xiong and R. Romanofsky⁶⁷ © IEEE 2005).

$[-\Delta\phi_{\max}, \Delta\phi_{\max}]$. It was shown in that the averaged effect of phase error is to introduce an amplitude loss of

$$L_{\Delta\phi} = \frac{\sin \Delta\phi_{\max}}{\Delta\phi_{\max}} \quad \text{or} \quad L_{\Delta\phi} (\text{dB}) = 20 \log \left| \frac{\sin \Delta\phi_{\max}}{\Delta\phi_{\max}} \right| \quad (21-21)$$

and there is no net phase shift of the composite signal due to the phase errors. For a maximum phase error of $\pi/8$, the loss is only 0.224 dB. It can be seen that the phase error loss is a sinc function of $\Delta\phi_{\max}$ and it can be shown that the maximum phase error must be less than 47° to limit the loss within 1 dB.

The real composite signal will have not only amplitude loss, but possibly also a small net phase error. The small net phase error could have a degrading impact on the BER performance of the high-order PSK schemes and QAM schemes, since the phase differences between constellation points are very small.

Effect of Finite Response Time on Beam Evolution

Different types of phase shifters will have different transient response times. Table 21-2 in the next section provides a qualitative assessment. When the array beam is switched from one direction to another, the finite response time of the phase shifters causes the beam pattern to evolve. This effect is in addition to the ISI and phase error described earlier. The problem was analyzed by incorporating the phase transient in the array factor. For the 925 element array investigated there, which assumed a quite long 0.3 ms average phase shifter response time, the directivity degraded by as much as 1.9 dB during position updates.

Experimental results were reported⁶⁸⁻⁶⁹ that would, of course, include all the effects described herein. Clearly, electrically large phased arrays communicating wide-band signals with high order modulation formats must be designed to compensate for these effects. A set of BPSK BER curves for various conditions is shown in Figure 21-16 (transient effect excluded). The term “without antenna” means that the antenna is ideal—no phase errors and no ISI.

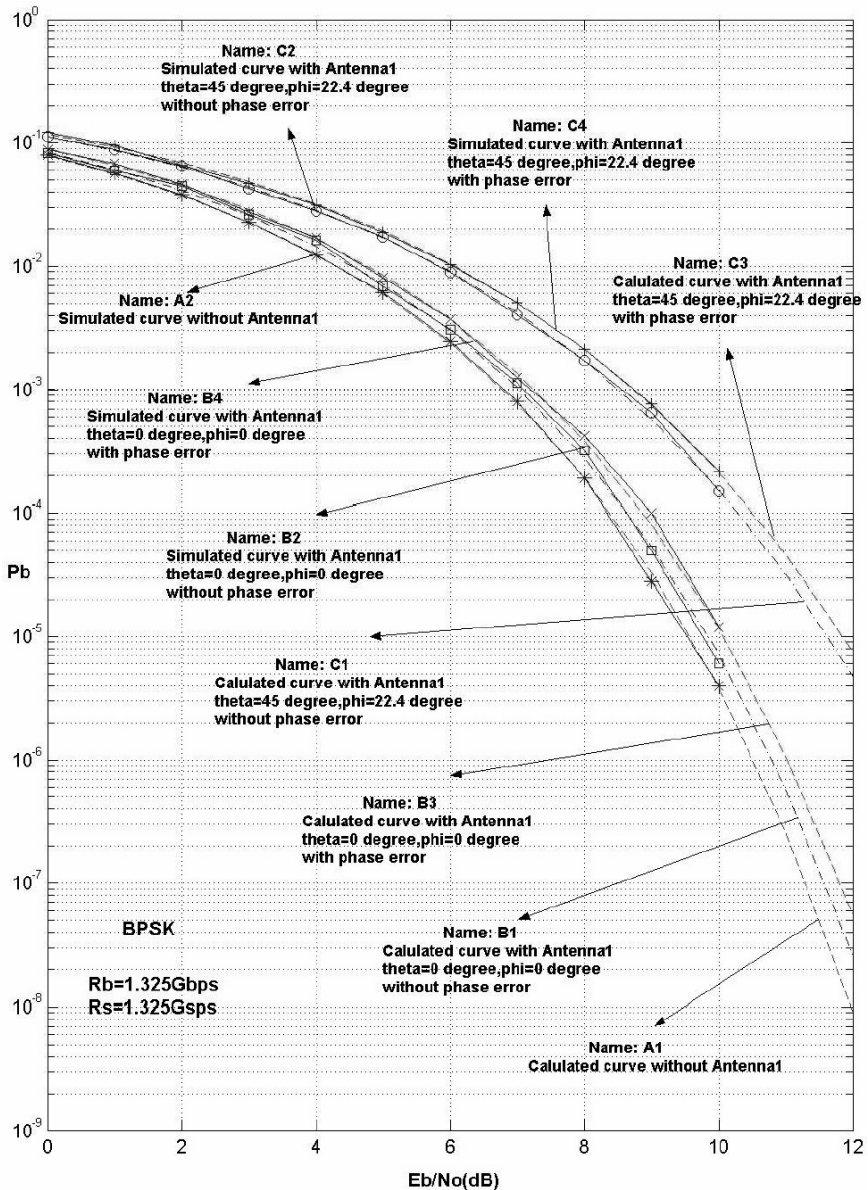


FIGURE 21-16 BER curves for BPSK (after F. Xiong and R. Romanofsky⁶⁷ © IEEE 2005)

TABLE 21-2 Comparison of Phase Shifter Technologies

Type Feature	Ferroelectric	Semiconductor/ MMIC	Ferrite	MEMS
Cost	Low	Expensive	Very expensive	Low
Reliability	Good after 10^6 0-40 V/ μm bias cycles (more tests required)	Very good (if properly packaged)	Excellent	Good after several billion cycles (more tests required)
Power handling	Good, >1W	Very good, tens of watts	Very high (kW)	Low Power, <50 mW for high reliability
Switching speed	Intrinsically fast ($\approx\text{ns}$), controller limited if high voltage	Fast at low power ($<10^{-9}\text{s}$)	Slow (inductance) 10 to 100 μs	Slow (mechanical) 10 to 100 μs
Radiation tolerance	Excellent	Poor (good if radiation hardened)	Excellent	Excellent (mechanical; no solid state junctions involved)
DC power consumption	Low ($<<1 \mu\text{A}$ current) μW	Low ($<10 \text{ mW}$ for diodes; negligible for FETs)	High (large current) $\approx 10 \text{ W}$ ($\approx 1 \text{ W}$ if latching)	negligible
RF loss	$\sim 5 \text{ dB}/36^\circ@$ K-band	$\sim 2 \text{ dB/bit}@$ Ka-band = 8 dB	$<1 \text{ dB} /$ $360^\circ@ \text{X-band}$	$\sim 2.3 \text{ dB}/337.5^\circ@$ Ka-band
Size	Very small if parallel plate varactor (mm^2)	Small ($\approx 10 \text{ mm}^2$ at Ka-band)	Large (wave-guide, bias coil)	Small (comparable to MMIC)
Linearity	IMD intersect +60 dBm	IMD intersect +35 to +40 dBm		IMD intersect +80 dBm

21.9 SUMMARY

This chapter reviewed basic phase shifter theory and described some modern phase shifter devices. Key developments of the past decade or so include devices based on GaAs MMIC, MEMS, and thin ferroelectric films. Applications of superconductivity were also briefly described. Some interesting effects are expected to manifest themselves as the ratio of signaling bandwidth to carrier frequency decreases, especially with regard to electrically large phased arrays. An important aspect of phase shifter performance is power handling capability. The reader is referred to the reference section⁷⁰⁻⁷³ for information on this topic. A summary of phase shifter characteristics is presented in Table 21-2.

REFERENCES

1. E. Brookner, "Major Advances in Phased Arrays: Part I," *Microwave Journal* (May 1997): 288-293.
2. E. Brookner, "Major Advances in Phased Arrays: Part II," *Microwave Journal* (June 1997): 84-92.

3. L. Eriksson and S. Broden, "High Performance Automotive Radar," *Microwave Journal* (October 1996): 24–38.
4. A. Stove, "Automobile Radar," *Applied Microwave Magazine* (Spring 1993) 102–115.
5. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes," *IEEE Int. Conf. Comm., Conf. Record*, vol. 2 (May 1993): 1064–1070.
6. J. Budinger et al, "Direct Data Distribution from Low Earth Orbit," *IEEE Int'l. Conf. Comm.*, (June 1997): 662–673.
7. C. A. Jensen, J. D. Terry, and M. Vanderaar, "The Implications of Encoder/Modulator/Phased Array Designs for Future Broadband LEO Communications," *SPIE Proc.*, vol. 3232 (1997): 61–72.
8. S. Ohmori, S. Taira, and M. Austin, *J. Comm. Res. Lab.*, 38 no. 2 (1991): 217.
9. D. Collier, "Optimizing LNAs for Use in Phased Arrays," *Microwave Systems News* (April 1990): 37–45.
10. L. Dunleavy, "GaAs MMICs Perform in Phase Shifters," *Microwaves & RF* (April 1984): 49–52.
11. J. White, *Microwave Semiconductor Engineering* (New York: Van Nostrand Reinhold Co., 1982).
12. G. Bartolucci, F. Giannini, and E. Limiti, "On the Generalized Loaded-Line Phase Shifter," *IEEE Int'l Microwave and Optoelectronics Conf. Proc.* (1995): 554–558.
13. J. Geddes et al, "Characteristics of 30 GHz MMIC Receivers for Satellite Feed Array Application," *GaAs IC Symposium Digest* (1987).
14. S. Ohmori, S. Taira, and M. Austin, "Beam Scanning Error of Phased Array Antenna," *Journal of Communications Research Laboratory*, vol. 38, no. 2 (July 1991) 217–222.
15. Z. Wang et al, "Single-Chip 4 Bit 35 GHz Phase-Shifting receiver with a Gb/s Digital Interface Circuitry," *GaAs IC Symposium* (1995): 234–237.
16. K. Bhasin et al, "Control of a GaAs Monolithic Ka-Band Phase Shifter Using a High-Speed Optical Interconnect," *IEEE Trans. MTT*, 38, no. 5 (May 1990): 686–688.
17. S. Rossek and C. Free, "Optically Controlled Microwave Switching and Phase Shifting Using GaAs FET's," *IEEE Micro. Guided Wave Letter*, vol. 5, no. 3 (March 1995): 8–83.
18. R. Romanofsky, "On the Relationship Between Schottky Barrier Capacitance and Mixer Performance at Cryogenic Temperatures," *IEEE Micro. Guided Wave Letter*, vol. 6, no. 8 (August 1996): 286–288.
19. W. McLevige and V. Sokolov, "Resonated GaAs FET Devices for Microwave Switching," *IEEE Trans. Elect. Dev.*, vol. ED-28, no. 2 (February 1981): 198–204.
20. V. Sokolov et al, "A Ka-band GaAs Monolithic Phase Shifter," *IEEE Trans. Microwave Theory Tech*, vol. MTT-31, no. 12 (December 1983): 1077–1082.
21. K. Kurokawa and W. Schlosser, "Quality Factor of Switching Diodes for Digital Modulation," *Proc. IEEE*, vol. 38 (January 1970): 180–181.
22. H. Atwater and R. Sudbury, "Use of Switching Q in the Design of FET Microwave Switches," *IEEE MTT-S Int. Microwave Symposium Digest* (June 1981): 370–372.
23. M. Schindler and A. Morris, "DC-40 GHz and 20-40 GHz MMIC SPDT Switches," *IEEE Trans. Micro. Theory Tech*, vol. MTT-35, no. 12 (December 1987): 1486–1492.
24. A. Tagantsev, "Mechanisms of Dielectric Loss in Microwave Materials," *Materials. Research Society Proceedings*, 603 (2000): 221–232.
25. R. Katiyar et al, "Investigations on Sol-Gel Derived $\text{Ba}_{0.5}\text{Sr}_{0.5}\text{Ti}_{1-\delta}\text{Mn}_{\delta}\text{O}_3$ Thin Films for Phase Shifter Applications," *Materials. Research Society Proceedings*, vol. 720 (2002): 3–14.
26. H. Wu and F. Barnes, "Doped $\text{Ba}_{0.6}\text{Sr}_{0.4}\text{TiO}_3$ Thin Films for Microwave Device Applications at Room Temperature," *Integrated Ferroelectrics*, vol. 22 (1998): 291–305.
27. N. Navi, J. Horowitz, H. Wu, and S. Qadi, "Structure-Property Relationships in W Doped $(\text{Ba,Sr})\text{TiO}_3$ Thin Films Deposited by Pulsed Laser Deposition on (001) MgO ," *Materials. Research Society Proceedings*, vol. 720 (2002): 41–48.
28. V. Varadan et al, "Ceramic Phase Shifters for Electronically Steerable Antenna Systems," *Microwave Journal* (January 1992): 116–127.
29. V. Varadan et al, "A Novel Microwave Planar Phase Shifter," *Microwave Journal* (April 1995): 244–253.

30. F. DeFlaviis, N. Alexopoulos, and O. Stafsudd, "Planar Microwave Integrated Phase Shifter Design with High Purity Ferroelectric Material," *IEEE Trans. MTT*, vol. 45, no. 6 (June 1997): 963–969.
31. O. Vendik, I. Mironenko, and L. Ter-Martirosyan, "Superconductors Spur Applications of Ferroelectric Films," *Microwaves & RF* (July 1994): 67–70.
32. J. Rao, D. Patel, and L. Sengupta, "Phased Array Antennas Based on Bulk Phase Shifting with Ferroelectrics," *Integrated Ferroelectrics*, vol. 18 (1998).
33. F. Van Keuls et al, "YBa₂Cu₃O_{7-δ}/Au/SrTiO₃/LaAlO₃ Thin Film Conductor/Ferroelectric Coupled Microstripline Phase Shifters for Phased Array Applications," *Applied Physics Letter*, vol. 71 (November 1997): 3075–3077.
34. F. Van Keuls et al, "Ku-Band Gold/ Ba_xSr_{1-x}TiO₃/LaAlO₃ Conductor/Thin Film Ferroelectric Microstripline Phase Shifter for Room Temperature Operation," *Microwave and Optical Tech. Letter*, vol. 20 (January 1999): 53–56.
35. R. Romanofsky, "Advances in Scanning Reflectarray Antennas Based on Thin Ferroelectric Film Phase Shifters," *Proc. IEEE, Special Issue on Technical Advances in Deep Space Communications and Tracking, to be published* (2007).
36. R. Romanofsky et al, "K-band Phased Array Antennas Based on Ba_{0.60}Sr_{0.40}TiO₃ Thin-Film Phase Shifters," *IEEE Trans. MTT*, vol. 48, no. (December 2000): 2504–2510.
37. R. Romanofsky, "Broadband, Low-Loss K- and Ka-Band Phase Shifters Based on Thin Ferroelectric Films," *IEEE MTT Symposium Workshop WMC*, Fort Worth, TX, June 7, 2004.
38. R. Romanofsky and A. Qureshi, "A Model for Ferroelectric Phase Shifters," *IEEE Trans. Mag.*, vol. 36, no. 5 (September 2000): 3491–3494.
39. G. M. Rebeiz, *RF MEMS Theory, Design, and Technology* (New York: John Wiley & Sons Inc., 2003).
40. B. Lakshminarayanan and T. Weller, "Design and Modeling of 4-bit Slow-Wave MEMS Phase Shifters," *IEEE Trans. MTT*, vol. 54, no. 1 (2006): 120–127.
41. N. Barker and G. Rebeiz, "Optimization of Distributed MEMS Phase Shifters," *IEEE MTT-S Digest* (1999): 299–302.
42. N. Barker and G. Rebeiz, "Distributed MEMS True-Time Delay Phase Shifters and Wide Band Switches," *IEEE Trans. MTT*, vol. 46, no. 11 (1998): 1881–1890.
43. M. Scardelletti, G. Ponchak, and N. Varaljay, "MEMS, Ka-Band Single-Pole Double-Throw (SPDT) Switch for Switched Line Phase Shifters," *IEEE International Symposium on Antennas and Propagation* (June 2002): 2–5.
44. M. Scardelletti et al, "RF MEMS Phase Shifters and Their Application in Phased Array Antennas," *IEEE Wireless and Microwave Technology Conference*, Clearwater, Florida, April 2005.
45. G. Rebeiz, G-L Tan, and J. Hayden, "RF MEMS Phase Shifters: Design and Applications," *IEEE Microwave Magazine* (June 2002): 72–82.
46. R. Romanofsky, "Slow-Wave Phase Shifters Based on Thin Ferroelectric Films, for Reflectarray Antennas," *IEEE MTT Symposium Workshop WMG* (June 2006): 121–137.
47. A. Nagra and R. York, "Distributed Analog Phase Shifters with Low Insertion Loss," *IEEE Trans. MTT*, vol. 47, no. 9 (September 1999): 1705–1711.
48. E. Erker et al, "Monolithic Ka-Band Phase Shifter Using Voltage Tunable BaSrTiO₃ Parallel Plate Capacitors," *IEEE Microwave and Guided Wave Letter*, vol. 10, no. 1 (January 2000): 10–12.
49. B. Acikel et al, "A New High Performance Phase Shifter Using Ba_xSr_{1-x}TiO₃ Thin Films," *IEEE Microwave and Wireless Comp. Letter*, vol. 12, no. 7 (July 2002): 237–239.
50. R. York, "BST Technology for RF Front Ends," *MTT Symposium Workshop WMG* (June 2006): 73–91.
51. G. Ponchak, "RF Transmission Lines on Silicon Substrates," *29th European Microwave Conference* (October 1999): 158–161.
52. G. F. Dionne et al, "Superconductivity for Improved Ferrite Devices," *Lincoln Laboratory Journal*, vol. 9, no. 1 (1996): 19–31.
53. W. Hord, "Microwave and Millimeter-Wave Ferrite Phase Shifters," *Microwave Journal, State of the Art Reference* (1989): 81–94.

54. T. Nelson et al, "Small Analog Stripline X-Band Ferrite Phase Shifter," *IEEE Trans. MTT Correspondence* (January 1970): 45–46.
55. D. Oates et al, "Superconductor Ferrite Phase Shifters and Circulators," *Proc. Appl. Superconductivity Conf.* (1996).
56. G. Dionne et al, "Ferrite-Superconductor Microwave Phase Shifters," *IEEE Trans. Mag.*, vol. 30, no. (November 1994): 4518–4520.
57. D. Durand et al, "The Distributed Josephson Inductance Phase Shifter," *IEEE Trans. Appl. Superconductivity*, vol. 2, no. 1 (March 1992): 33–38.
58. E. Track et al, "Investigation of an Electronically Tuned 100 GHz Superconducting Phase Shifter," *IEEE Trans. Mag.*, vol. 27, no. 2 (March 1991): 2700–2703.
59. J. Bednorz and K. Muller, "Possible High T_c Superconductivity in the Ba-La-Cu-O System," *Z. fur Phys.*, 64 (1986): 189.
60. Z-Y Shen, *High-Temperature Superconducting Microwave Circuits* (Norwood, MA: Artech House, 1994).
61. J. Martens et al, "A Reflective Switch Made of Tl-Ca-Ba-Cu-O for Signal Control Applications," *IEEE Microwave and Guided Wave Letter*, vol. 1, no. 10 (1991): 291–293.
62. Y. Tzeng et al, "High Temperature Superconductor Opening Switch," *Applied Physics Letter*, vol. 54, no. 10 (1989): 949–950.
63. W. Donaldson et al, "Interaction of Picosecond Optical Pulses with High-T_c Superconducting Films," *Applied Physics Letter*, vol. 54, no. 24 (1989): 2470–2472.
64. C. Jackson et al, "A High Temperature Superconductor Phase Shifter," *Microwave Journal* (December 1992): 72–78.
65. P. Mason and R. Gould, "Slow-Wave Structures Utilizing Superconducting Thin-Film Transmission Lines," *J. Appl. Phys.*, vol. 40., no. 5 (April 1969): 2039–2051.
66. W-L Cao et al, "Picosecond Superconductor Opening Switches," *IEEE Trans. Appl. Superconductivity*, vol. 3, no. 1 (1993): 2848–2851.
67. F. Xiong and R. Romanofsky, "Study of Behavior of Digital Modulations for Beam Steerable Reflectarray Antennas," *IEEE Trans. Ant. and Prop.*, vol. 53, no. 3 (March 2005): 1083–1096.
68. O. S. Sands, "Beam-Switch Transient effects in the RF Path of the ICAPA Receive Phased Array Antenna," *NASA Technical Memorandum TM2003-212588* (February 2002).
69. R. Reinhart et al, "Phased Array Antenna-Based System Degradation at Wide Scan Angles," *IEEE International Symposium on Phased Array Systems and Technology, Revolutionary Developments in Phased Arrays*, Boston, MA, October 2003.
70. S. K. Koul and B. Bhat, *Microwave and Millimeter Wave Phase Shifters: Vol. II Semiconductor and Delay Line Phase Shifters* (Boston: Artech House, 1991).
71. A. Kozyrev et al, "Response Time and Power Handling Capability of Tunable Microwave Devices Using Ferroelectric Films," *Integrated Ferroelectrics*, vol. 22 (1998).
72. M. Hines, "Fundamental Limitations in RF Switching and Phase Shifting Using Semiconductor Diodes," *Proc. IEEE* (June 1964): 697–708.
73. J. White, "Review of Semiconductor Microwave Phase Shifters," *Proc. IEEE*, vol. 56, no. 11, (November 1968): 1924–1930.

Chapter 22

Conformal and Low-Profile Arrays

Robert J. Mailloux

University of Massachusetts, Amherst

CONTENTS

22.1	INTRODUCTION.....	22-2
22.2	FUNDAMENTAL PRINCIPLES.....	22-3
22.3	ANTENNAS AND ARRAYS ON CYLINDERS.....	22-9
22.4	PATTERN SYNTHESIS FOR CYLINDRICAL SECTOR ARRAYS	22-20

22.1 INTRODUCTION

A number of applications require antenna arrays conforming to a nonplanar surface. Primary among these are requirements for scanning antennas on aircraft or missiles, where aerodynamic drag is reduced for a conformal flush-mounted geometry, and for arrays that conform to surfaces that provide some coverage advantage, for example, to a hemispherical surface for hemispherical-pattern coverage or to a cylinder for 360° coverage in the azimuth plane.

Conformal arrays can be grouped into two broad categories that have very different design goals and that present significantly different technological challenges. Arrays in the first category are small with respect to the radius of curvature of the conformed body, as depicted in Figure 22-1a. These proportions are representative of flush-mounted and low-profile aircraft arrays at superhigh-frequencies (SHFs) and above. These are relatively flat and behave nearly like planar arrays. The technological challenges in this case are to design an array to scan over very wide angles (for example, zenith to horizon) or to make the array so thin that it can conform to the outer surface of the aircraft without requiring large holes in the fuselage.

The second category of conformal arrays, depicted in Figure 22-1b, is the type more commonly regarded as conformal since the array is comparable or large with respect to the radius of curvature of the mounting body. Such arrays include cylindrical arrays developed to provide radiation from spinning missiles and for 360° coverage of ground-based radars. Conical and spherical arrays have also been developed for ground-based radars. Conical and spherical arrays have also been developed for ground, airborne, and missile applications.

Among the special features that make it difficult to design arrays that are large or highly curved in this sense, the following are usually listed:^{1,2} (1) Array elements point in different directions, so it is often necessary to switch off those elements that radiate primarily away from the desired direction of radiation. (2) Conformal-array synthesis is very difficult because you cannot factor an element pattern out of the total radiation pattern. (3) Mutual coupling can be severe and difficult to analyze because of the extreme asymmetry of structures like cones and because of multiple coupling paths between elements (for example, the clockwise and counterclockwise paths between elements on a cylinder). Another phenomenon related to mutual coupling is the evidence of ripples on the element patterns of cylindrical arrays. (4) Cross-polarization effects arise because of the different pointing directions for elements on

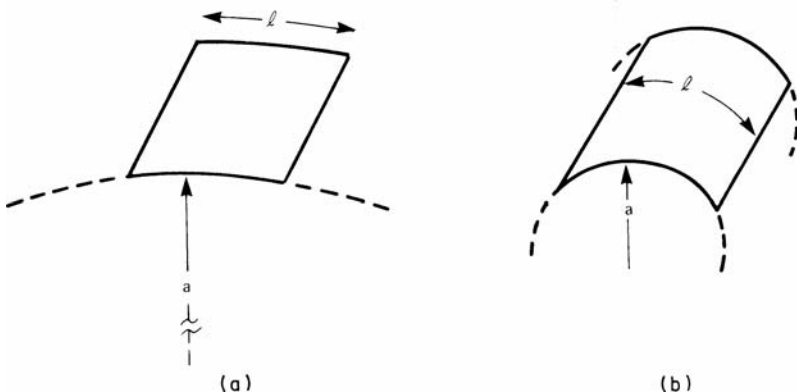


FIGURE 22-1 Conformal-array apertures: (a) aperture dimensions much less than local radius of curvature and (b) aperture dimensions comparable with local radius of curvature

curved surfaces, which cause the polarization vector projections to be nonaligned. (5) Even for a cylindrical array there is a need to use different collimating phase shifts in the azimuth plane for an array scanned in elevation because steering in azimuth and elevation planes is not separate.

22.2 FUNDAMENTAL PRINCIPLES

Properties of Isolated Elements on Finite Planes and Cylinders

Radiating elements behave differently according to the structure on which they are mounted. Examples of elementary slot antennas radiating from circular cylinders show significant pattern changes at angles near or below the horizon for a slot located at the top of the cylinder.

Figure 22-2 shows the radiated power pattern in the upper hemisphere ($\theta \leq 90^\circ$) for an infinitesimal slot in a cylinder of radius a . Using the results of Pathak and Kouyoumjian,³ you can express the radiation pattern of elementary slots on cylinders in terms of three distinct regions, depending upon whether the point of observation is in the upper hemisphere (called the *illuminated region*), in the lower hemisphere (called the *shadow region*), or in the vicinity of the horizon (called the *transition region*). The angular extent of the transition region is of the order of $(k_0 a)^{-1/3}$ on each side of the shadow boundary. Figure 22-2 illustrates that within the illuminated region the circumferentially polarized radiation is nearly constant for an infinitesimal element but that the axially polarized radiation has a $\cos \theta$ element pattern. Moreover, within the transition region these patterns are modified so that the circumferentially polarized component drops to a value of about 0.7 in field strength, or -3.2 dB as compared with its value of zenith, while for axial polarization the horizon radiation is not zero but approximately

$$0.4 \left(\frac{2}{k_0 a} \right)^{1/3}$$

Thus for a thin slot with circumferential polarization, the horizon gain is independent of the size of the cylinder, but for an axially polarized slot, the horizon gain varies directly as a function of $(k_0 a)^{-1/3}$. For cylinders of radius approaching 50λ , the horizontally polarized power at the horizon is thus approximately 23 dB below zenith gain.

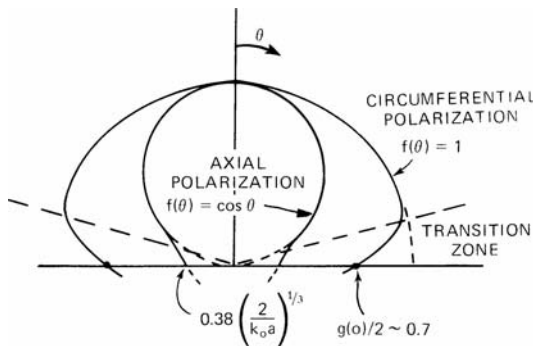


FIGURE 22-2 Approximate pattern of slot on cylinder of radius a

Pattern Coverage Limitations for Conformal Arrays

In addition to the diffraction effects observed in the preceding section, purely geometrical effects play a major part in determining the pattern characteristics of conformal arrays. Figure 22-3 shows a conical array of slots from several angles and illustrates the polarization distortion that occurs for an array near the tip of a cone. This figure, due to Kummer,¹ shows three views of a conical array with radiators located along the generatrices and oriented to give linear polarization in the axial direction. Clearly this array radiates a different polarization for each chosen angle of radiation.

A second issue raised by the picture of the conical array is that all of the elements have different element patterns because each is a different distance from the tip and each is at a different azimuth angle. Even if the elements near the tip are left out (since they are few and so have relatively little gain), only a small number of radiators have element patterns that point in the required direction. It is therefore usually necessary to devise a sophisticated network to switch an illuminated sector to the proper location on the conformal array for efficient radiation. Such networks now exist for cylindrical arrays, but the problem is more severe for cones, spheres, and more general surfaces. In addition, because of the polarization rotation, it may be necessary to align the element polarization in order to radiate efficiently. Finally, it is difficult to maintain good monopulse nulls for directions in which there is not even symmetry.

Since the elements may point in different directions, there is a size limit for an array that can be wrapped around a conformal body without sacrificing efficiency. Figure 22-4a shows the radiation pattern of several cylindrical arrays computed as a function of the angle subtended by the array. As pointed out by Hessel,⁴ for an array of $k_0 a = 86$, with 0.65λ separation between elements, doubling the number of excited elements in the array (from a 60° to a 120° arc) increases the array gain by only about 1 dB because the edge elements do not radiate efficiently toward the beam peak. The large sidelobes near $\phi = 100^\circ$ are associated with grating lobes of the elements grouped at the array ends, which are necessarily scanned to very wide angles.

Closer element spacing can relieve the grating-lobe problem and can broaden the element patterns, as will be described in Section 22.3, but Figure 22-4b shows that the gain restrictions are fundamental limitations. This figure shows the gain of a cylindrical array

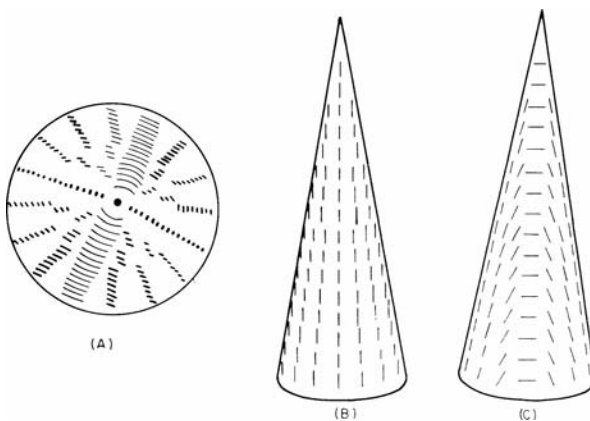


FIGURE 22-3 Conical array: (a) view from axial direction, (b) side view, and (c) side view rotated 90° from b (after Kummer¹ © IEEE 1974)

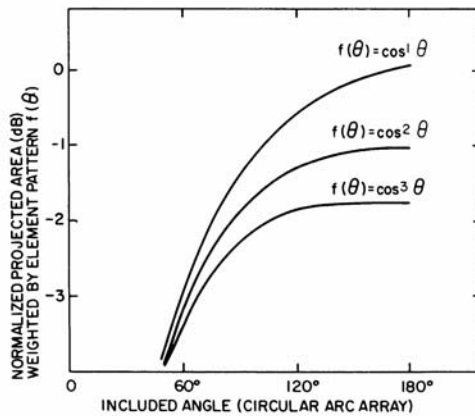
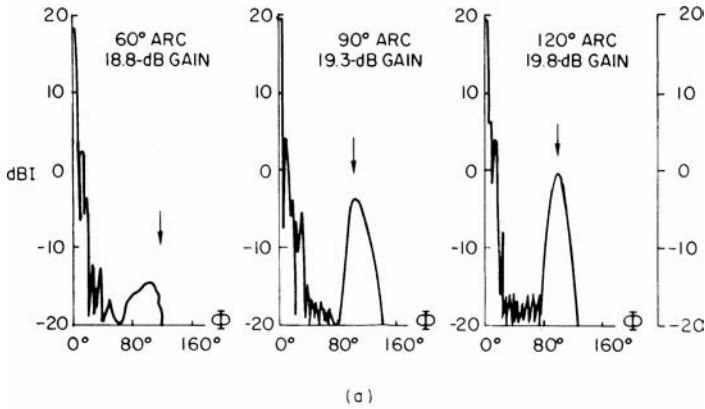


FIGURE 22-4 Radiation characteristics of arc arrays: (a) maximum-gain E -plane patterns for arc array of $k_0 a = 86$, $b/\lambda = 0.65$ (after Hessel⁴ © Artech House 1972) and (b) normalized projected area weighted by element patterns

with assumed power patterns $\cos^1 \theta$, $\cos^2 \theta$, and $\cos^3 \theta$. It is also assumed that the element spacing approaches zero (the continuous aperture case) and the gain, normalized to the gain of the projected area of the cylindrical section, is the integral of the \cos^n functions up to the maximum θ value. This is a good approximation for radii greater than about 5 wavelengths, and for element spacing up to a half wavelength. Additionally, the circular arc arrays require more elements than a planar array with the same projected area, so if the normalized gain were normalized to the number of elements to fill the sector, then the comparison with a planar array is made worse. The circular sector with included angle $2\theta_0$ requires about $\theta_0/\sin(2\theta_0)$ more elements than the projected aperture [a factor of 1.1 (0.41 dB)] for a 90° sector, and 1.6 (2.04 dB) for a 180° sector. These two limitations show that there is little advantage to exciting much more than a 90° section of the cylinder.

A second fundamental result is obtainable from the same very simplified analysis. For a four-faced square array with perimeter equal to the circular-array circumference, the average gain over a 90° scan sector, normalized again to the cylinder projected area, is approximately -1.5 dB for a $\cos \theta$ element pattern, and -1.9 dB for the $\cos \theta^2$ element pattern. Figure 22-4b shows that these are essentially the same as the gain of a 90° arc circular array with the same element pattern. A similar result is quoted by Hessel.⁴

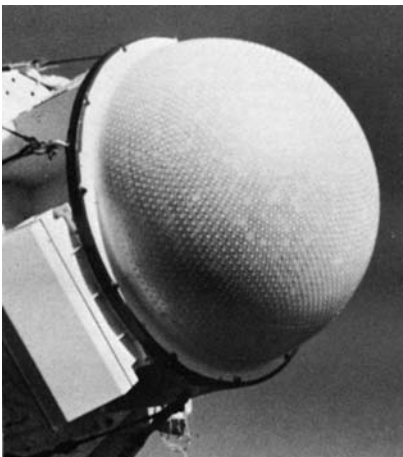
The preceding results are derived for uniformly illuminated arrays. When arrays have severe illumination tapers, as for example in low sidelobe arrays, it is possible to occupy larger circular arcs without reduced efficiency because of the low edge illumination.

Technology of Conformal Arrays

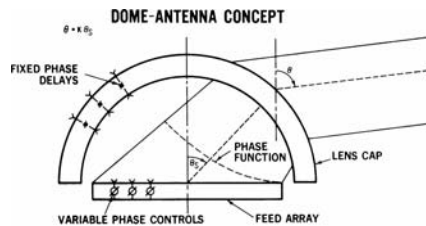
The dramatic illustration chosen by Kummer¹ (see Figure 22-3) points out the difficulty of synthesis on conical and more generalized surfaces that arises because such structures project differently into different angular regions. Nevertheless, cylinders, cones, and spheres are important surfaces for conformal array development because of their obvious application to missile and certain hemispheric scanning structures.

The technology of conformal arrays is very different from that of planar arrays. Often conformal arrays must be flush mounted or low profile, and structurally integrated with the platform with little intrusion. Thus, even for arrays that do not wrap around the body of the vehicle significantly, there is still a basic difference in the technology. The second major distinction is that the excitation of an array that is large with respect to the local radius of curvature needs to be commutated around the body, and this can be a significant burden. Commutators devised for cylindrical arrays are described in Section 22.3. Analog commutating networks for general conformal arrays require complex switching trees that are costly, lossy, and hard to implement, and that often require both amplitude and phase control.

One special type of space-fed commutating network has been developed for cylinders and spheres. The Dome antenna,^{5,6} shown in Figure 22-5 in its spherical version, uses the vertical projection of a spherical dome to achieve increased gain at low angles of elevation. The dome is a passive spherical lens made up of fixed phase shifters and fed by a conventional planar



(a)



(b)

FIGURE 22-5 Hemispherical dome array: (a) dome array and (b) dome-antenna concept (Courtesy of Sperry Corporation)

phased array that steers an illuminated spot to various portions of the lens. Fixed phase shifters in the lens are computed so as to convert the array scan direction θ into a factor K times that angle, and so as to obtain scan in excess of the planar array scan angle. The array phase shifter settings are determined to form the nonlinear phase progression required to scan the searchlight-type beam to various spots on the lens. Although the radiated beamwidth varies with scan angle, the dome can achieve scanning over sectors larger than a hemisphere and has achieved scan to $\pm 120^\circ$ from zenith. Steyskal et al⁶ give equations for the gain limits of a given circular cylindrical dome based upon allowable scan angles for the feed array and show that the ratio of the average gain to the broadside feed array gain is bounded by a parameter that depends upon the feed array scan limit.

In the general case where there is not enough volume to use a space feed, other means of commutating the illumination are needed to scan a large conformal array. There are several new technologies that can accomplish this. The first is the technology of solid state T/R modules to control the aperture illumination. In this case, you can control the distribution using resistive attenuators and linear amplifiers or by using amplifiers with variable gain to move the amplitude distribution around the array structure and so achieve high quality patterns, even for highly curved arrays.

Figure 22-6 shows⁷ a faceted array of dual polarized dipole elements conformal to a sphere and using six elements per facet. The array is intended for satellite communication from mobile units and has integrated transmit-receive electronics, using GaAs MMIC circuits.

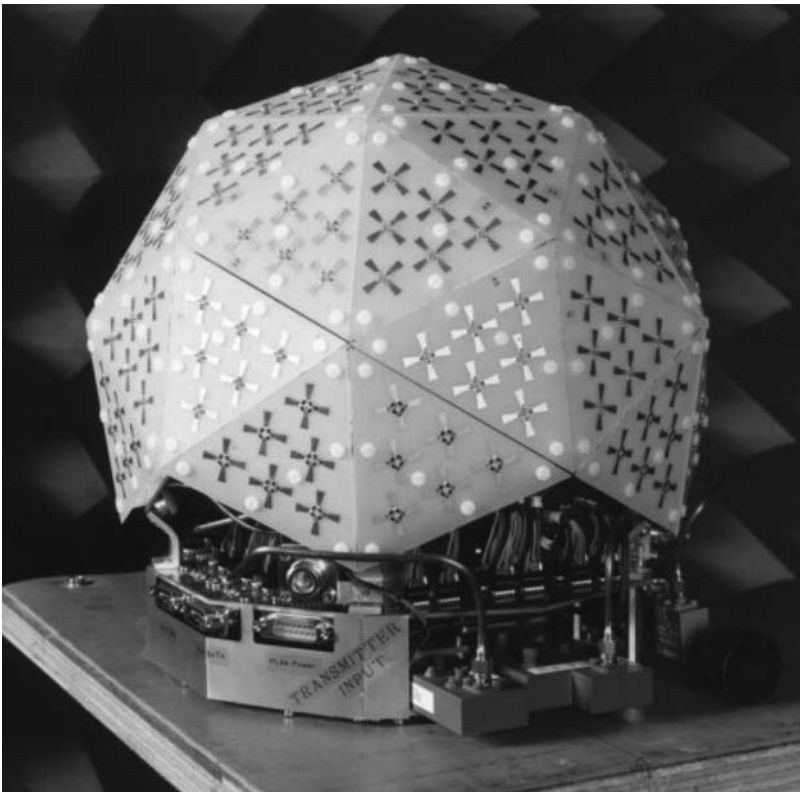


FIGURE 22-6 Faceted spherical array (Courtesy of Roke Manor Research Ltd., Roke Manor, UK)

The second approach is to implement full digital beamforming across the array. This approach, which can be used for transmit and receive, has thus far been implemented in a number of receive arrays that transmit using conventional analog methods. In the receiving array, the received signals are amplified, possibly down converted, and then digitized. Amplitude and phase weights for commutation are applied to the digitized signals. Though presently expensive, this technology is seen as the ultimate in performance. Three significant examples were presented at the 1996 Phased Array Symposium showing the application of digital beamforming to conformal receive arrays. The first⁸ was an array of 570 *X*-band active T/R modules that excite an array conformal to an elliptical cylinder. Measured low sidelobe receive patterns were below -38 dB, and additional measured patterns demonstrated the accurate synthesis of flat-topped sector patterns, cosecant squared patterns, and several patterns with anti-clutter nulls. In each case the patterns were synthesized using measured element patterns. A second example⁹ used an array of 576 T/R modules conformal to a doubly curved surface. Sample patterns confirmed about the same level of sidelobes as the previous study, and the ability to reconfigure the aperture to produce a variety of shaped patterns. A hemispherical array¹⁰ using the same T/R modules formed sidelobes at the -40 -dB level.

Another example of the power of these digital techniques for conformal arrays is the work of Steyskal,¹¹ who demonstrated a technique for forming frequency independent patterns with a circular cylindrical array. The symmetry property of a circular array of N elements leads to expansion of the fields in terms of orthogonal "phase modes,"¹² of signals that advance in phase with interelement phase $2\pi n/N$ for $1 \leq n \leq N$. This orthogonality corresponds to the beam port relations of an orthogonal beam (Butler) matrix, or a digital FFT. Steyskal's procedure modifies the mode excitation coefficients digitally and obtains a radiation pattern with angular dependence independent of frequency. Experimental data showed a sequence of three patterns at different frequencies (4, 5, and 6 GHz) with essentially the same patterns.

Methods of Analysis

The analysis of conformal antennas and arrays is undertaken with a variety of methods, depending upon whether the antenna or array dimensions are small with respect to the platform radius of curvature or whether the platform itself is large or small with respect to the operating wavelength. Full numerical solutions (moment method, finite element, or finite difference methods) are still not practically applied for electrically large host bodies, and to date, hybrid methods have proven more useful for these. Array analysis involves an additional complication, in that the interaction between elements, or mutual coupling, must be included in the solution for the fields and currents. This analysis is also very dependent upon the host body dimensions and relative curvature.

Historically, the first methods to find utility were based on integral equation solutions. These pertained to structures like spheres, cones, and infinite cylinders for which full Green's functions exist. They have become very practical to solve using the numerical method of moments (MOM) algorithms.¹³⁻¹⁶ The need to analyze antennas on dielectric substrates has led to the use of spectral methods that have become practical because of the increased computing power now available.¹⁷⁻¹⁹ These conformal structures are analyzed by MOM using some form of a spectral Green's function and matching to boundary conditions either in the spectral domain or after transforming back to the spatial domain.

Descardecì and Giarola²⁰ combined the cavity method with the use of the dyadic Green's function for a perfectly conducting cone, to obtain impedance and radiation patterns for microstrip patch antennas on cones. They assumed that the radius of curvature and cone tip distance was large compared to the wavelength. A related approach, but one not requiring a

geometry with an exact Green's function, allowed Jin et al²¹ to evaluate the radiation from microstrip elements and arrays on cylindrical bodies of arbitrary cross section. In this case the cavity problem is solved with the finite element technique, while the external fields are obtained using a method of moments solution based on reciprocity.

Kildal and Sanford²² and Sipus et al²³ have developed software using the spectral domain method of moments techniques for cylinders of arbitrary cross section, including multiple dielectric layers. They have demonstrated the technique for spherical structures as well. More general shapes have been studied using the free space Green's function and the method of moments using combined field integral equations as well as magnetic and electric field integral equations where appropriate. The Fast Illinois Solver Code (FISC) was developed to treat large problems using the iterative conjugate gradient method and fast multiple techniques. These methods are discussed in several references.^{24,25}

Elements and arrays on generalized surfaces have also been analyzed using the finite difference time domain (FDTD)^{26,27} and finite element (FEM)^{28,29} methods. These methods both offer the flexibility to treat elements on arbitrary three-dimensional surfaces and including dielectric volumes.

Analytical results for antennas on electrically larger bodies have been obtained using a combination of MOM or finite element techniques³⁰ plus quasi-optical techniques. Early work had its origins in the use of the geometrical theory of diffraction (GTD) for simply computing single-term mutual coupling between elements on general convex cylinders and on cones^{31,32} and grew into methods that used GTD to compute the local fields for MOM solution.

More recent work^{33,34} extends the utility of the hybrid methods using new Green's functions developed from the uniform theory of diffraction (UTD). Demirdag and Rojas³⁴ used these Green's functions for evaluation of the mutual coupling between elements on a uniformly coated, perfectly conducting, but otherwise arbitrarily shaped convex surface. The procedure uses generalized high frequency UTD solutions for circular cylinder and sphere to develop solutions for the arbitrarily shaped convex surface via the local EM wave propagation at high frequencies. The demonstrated example evaluated coupling between current elements on a dielectric coated circular cylinder with inner radius 3λ and a coating 0.06λ thick. The coating dielectric constant ϵ_r was 3.25. These results compared well against the exact eigen function expansion for elements greater than 0.5λ apart.

Persson and Josefsson,³⁵ also using UTD-based techniques, investigated the mutual coupling of apertures and arrays of apertures in circular cylinders. The accuracy of this approach was such as to demonstrate agreement with measured data down to the -60- to -80-dB levels. These and other significant results are also recorded in the book by Josefsson and Persson.⁷ Figure 22-7 shows the cylindrical array used to evaluate coupling and array performance. This array has 54 azimuth polarized elements arranged in three rows conformal to a cylinder. The array is 600 mm in diameter. The authors measured mutual coupling and other data.

22.3 ANTENNAS AND ARRAYS ON CYLINDERS

Wide Angle Coverage with Fuselage-Mounted Arrays

One of the major applications for vehicle arrays is for aircraft-to-satellite communications, for which the antenna system is often required to provide coverage throughout the entire hemisphere, usually with circular polarization and including substantial gain at the horizon. Figure 22-8 shows several fully electronically steered arrays, two of which are conformal, while the third is a low-profile, four-faced array.

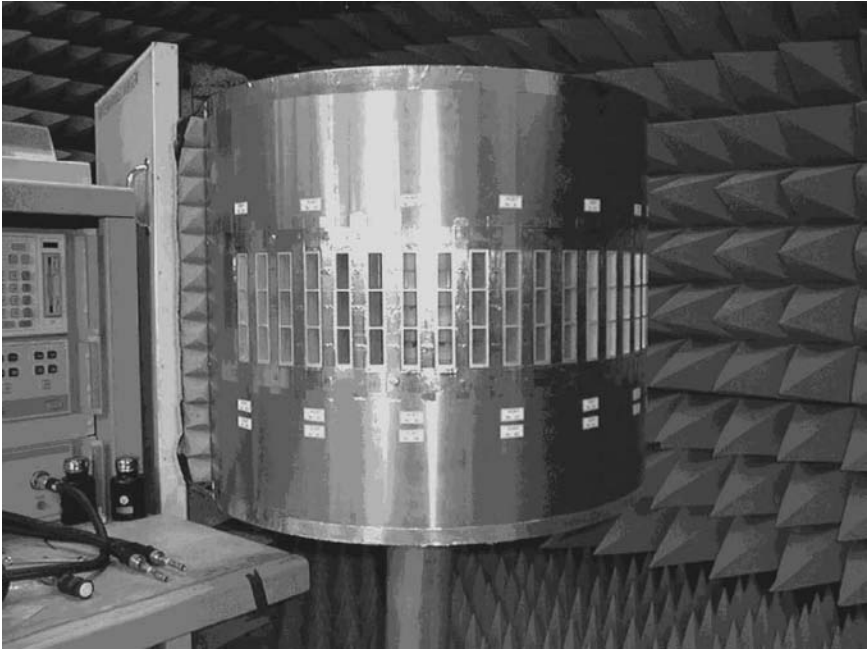


FIGURE 22-7 Array of 54 azimuthally polarized waveguide apertures (*Courtesy of Ericsson Microwave Systems, AB, Goteborg, Sweden*)

The use of a single array (see Figure 22-8a) is very appealing provided that gain-contour requirements can be met. Unfortunately, this means that the array must scan from zenith to horizon in all planes. The following illustrates some of the gain limitations for an array scanned from zenith to the horizon.

In an infinite two-dimensional array with no grating lobes, the pattern directivity for a perfectly matched aperture varies like $\cos \theta$. If the array mismatch is not corrected as a function of scan angle, the gain will vary according to the expression

$$(|1 - |\Gamma|^2|) \cos \theta$$

For finite arrays the beamwidth in the plane of scan varies much like $\sec \theta$ except near end fire. The elevation beamwidth δ for an array scanned to 90° (horizon) is given³⁶ approximately by the equation below for column and two-dimensional arrays:

$$\delta = \sqrt{\frac{4\pi}{C} \left(\frac{\lambda}{L_1} \right)}$$

Here L_1 is the length of the array in the end-fire direction, and C is the constant that varies between 3.5 and 7, depending upon phase velocity and the array distribution. If you

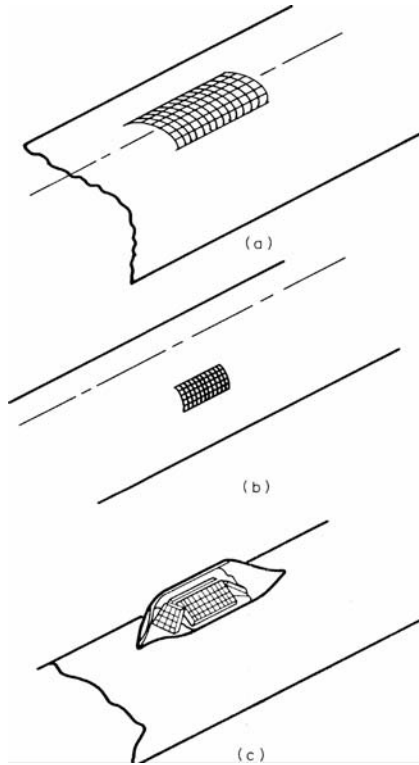


FIGURE 22-8 Conformal and low-profile arrays for hemispherical coverage: (a) large array on top of fuselage, (b) array conformal to side of fuselage, and (c) four-faced array

assume an azimuth beamwidth of approximately λ/L_2 is the array dimension perpendicular to the plane of scan, the directive gain at end fire is approximately

$$D = 4\pi \frac{L_2}{\lambda} \frac{1}{\delta} = \sqrt{4\pi C} \left(\frac{L_2}{\lambda} \right) \sqrt{\frac{L_1}{\lambda}}$$

Since the directivity is multiplied by four for an array of slots over a perfect ground plane (and the beamwidth is halved), the ratio of directive gain at end fire to directive gain at broadside (with perfectly conducting ground) is approximately

$$\frac{D}{D_0} = \sqrt{\frac{C}{\pi}} \sqrt{\frac{\lambda}{L_1}}$$

The result for an array of elements other than slots is different by a constant.

This equation shows that for any given broadside directive gain D_0 , it is advantageous to reduce L_1 (the array projection in the end-fire direction) as much as possible,

for that minimizes scan loss. It also indicates that larger arrays undergo increased scan loss. For example, a square array with 20-dB gain at broadside should exhibit a directivity degradation of 1 dB at end fire, but for a 30-dB square array a 4.5-dB falloff is expected.

Unfortunately, this decrease in directivity is only one of the factors tending to make scanning to end fire inefficient. Other major factors are diffraction and scattering due to the vehicle on which the array is mounted (see Section 22.1), element-pattern narrowing, and array mismatch due to mutual coupling.

Other options for hemispherical coverage shown in Figure 22-8 include the use of multiple conformal arrays (see Figure 22-8*b*), one on each side of the vehicle to provide good coverage except in conical regions near the vehicle nose and tail, or a four-faced array (see Figure 22-8*c*), perhaps with adequate coverage throughout the hemisphere with a relatively streamlined geometry. Multiple planar arrays are often used for ground-based-radar coverage, and Corey³⁷ gives a convenient graphical technique for determining the vertical face tilt and element packing geometry for planar antennas with given scan limits.

Figure 22-9 shows contours of constant projected area for any of the configurations of Figure 22-8, using the coordinate relationships that follow, for a planar array tilted at angle δ from the vertical:

$$\Theta = \cos^{-1} [v \cos \delta + c \sin \delta]$$

$$\phi = \tan^{-1} \left[\frac{u}{-v \sin \delta + c \cos \delta} \right]$$

where $u = \sin \theta \cos \phi$, $v = \sin \theta \sin \phi$, and $c = \cos \theta$ is the array projection factor. In general, the array projected area does not equate directly to relative gain because of array mismatch,

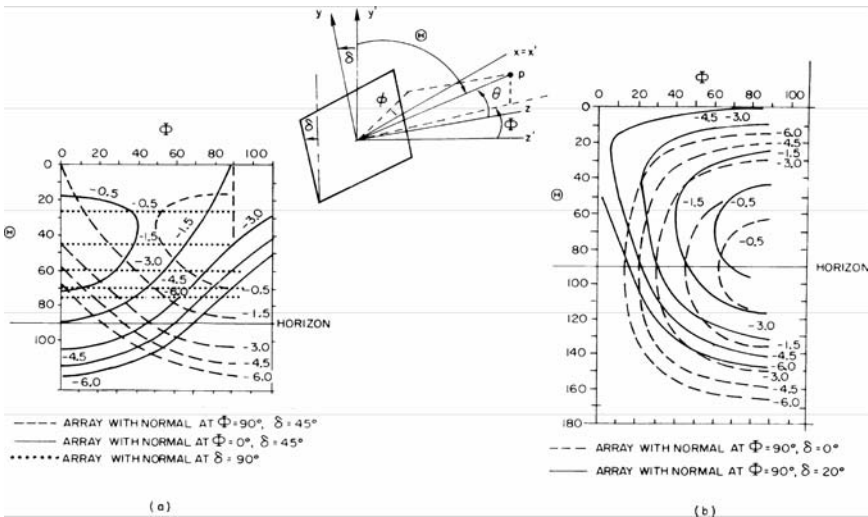


FIGURE 22-9 Coverage diagrams for a variety of tilted planar-array faces: (a) arrays tilted to $\delta = 45^\circ$ with array normals at $\phi = 0^\circ$ and 90° and array with normal at zenith ($\delta = 90^\circ$) and (b) arrays with normal at $\phi = 90^\circ$ and at $\delta = 0^\circ$ and 20°

element-pattern variation, and polarization effects, but it does provide a useful upper bound. This optimum-gain contour, normalized to peak, is given by

$$\frac{P}{P_0} = 10 \log c$$

Figure 22-9a shows three contour plots of this ratio: two for arrays mounted at $\delta = 45^\circ$ and one with $\delta = 90^\circ$. The dotted line is the gain contour for an array with its boresight at zenith ($\delta = 90^\circ$) and corresponds approximately to Figure 22-8a. The solid curve corresponds to an array located as in Figure 22-8b or a side face of the four-faced array of Figure 22-8c mounted with $\delta = 45^\circ$, $\Phi = 0^\circ$. The dashed curve is for an array with its boresight in the $\Phi = 90^\circ$ plane and corresponds to the end face of Figure 22-8c. Combining the solid and dashed curves, as is done for the four-faced array, indicates that an array with these four faces would cover the upper hemisphere with only a 3-dB projected area falloff near $(\Theta, \Phi) = (90^\circ, 45^\circ)$, while the array with $\delta = 90^\circ$ would have less gain for all $\Theta > 60^\circ$, and thus has its lowest gain over the large area near the horizon.

A more streamlined four-faced array is obtained by reducing the front and rear faces to about one-half of the size of the side faces. In this case it is convenient to mount these arrays with their boresight direction closer to the horizon because the larger side faces provide sufficient gain at the higher elevation angles in the $\Phi = 90^\circ$ or 180° planes. Figure 22-9 shows two gain contours for full-size arrays with $\delta = 20^\circ$ (solid) and $\delta = 0^\circ$ (dashed) for a front-face $\Phi = 90^\circ$. By evaluating the coverage contours in conjunction with those of the solid curve from Figure 22-9a (side face), you can compare the available coverage for a four-faced array with a variety of array sizes and elevation angles. If the front and rear faces are selected to be smaller than the side faces, the corresponding number of decibels (for example, 3 dB for one-half of the area) must be subtracted from the values on the contour lines.

As an example, you can show that, with half-area front and rear faces and using $\delta = 20^\circ$ for these end faces, the array system can scan throughout the hemisphere with a minimum projected area of approximately -4.5 dB with respect to the maximum.

Figure 22-9 can also be used with $\cos^2 \theta$ element patterns by doubling all the decibel ratios given in the figures or with any other assumed element patterns that are independent of Φ by making the appropriate change of scale.

Arrays mounted above a cylindrical surface are subject to pattern deterioration owing to multipath scattering from the surface. Pattern calculations for a four-faced array (see Figure 22-8c) have been carried out by using geometrical optics³⁸ using expressions from the work of Kouyoumjian and Pathak.³⁹ Figure 22-10 shows the computed gain for both polarizations by using a small array mounted at 45° from the top of a large cylinder. The indicated pattern ripples are the result of a specular multipath from the cylinder surface, and the figure shows that in this particular instance these ripples do not materially alter the elevation-plane coverage of the array.

Satellite communication from airborne terminals at EHF also requires near hemispherical coverage. Again the cost of two-dimensional scanning has stimulated development of alternative electromechanical arrays. Figures 22-11a and b show a sketch of one concept⁴⁰ for providing elevation electronic scan from zenith to 20° from the horizon, and electromechanical scan in the azimuth plane. The antenna array is tilted 35° from the horizontal and mounted below the surface of the cylinder. It is mounted off-center in a circular frame, and the whole system is rotated to provide 360° of azimuth coverage. The array requires an oversize hole in the fuselage in order to provide coverage to the 20° angle. Figure 22-11b shows a brassboard demonstration array by Hazeltine Corporation following this concept. The array receives at 20 GHz and transmits at 40 GHz. The brassboard model shown in the figure obviously has significant intrusion (over two feet), but a later version of this array

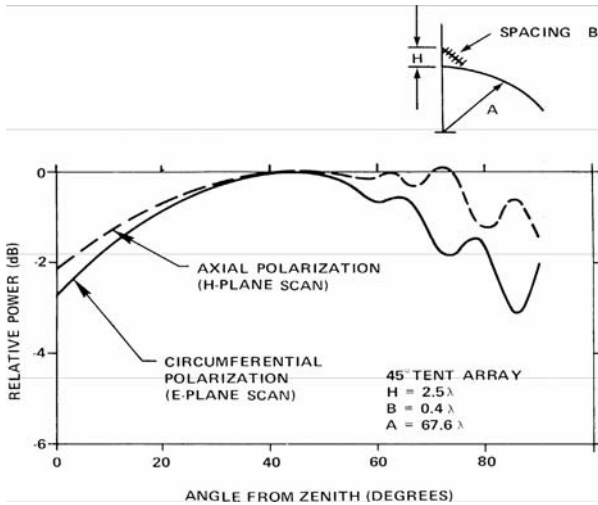
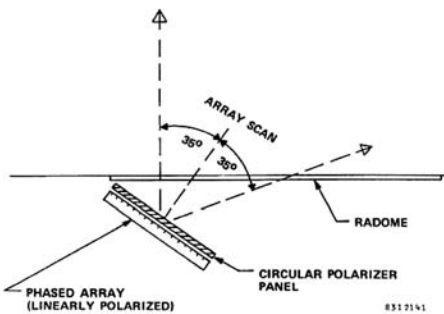


FIGURE 22-10 Multipath effects on a tilted four-faced array

was designed to be about 11 inches deep. The 6- by 14-inch aperture is a unique design incorporating both frequencies in an interlaced configuration of slots. Key features of the design include parallel plate feeds, interlaced linearly polarized slot elements, and a dual frequency meander line polarizer with good characteristics throughout the $\pm 35^\circ$ elevation scan, and a dual-frequency radome with 0–70° scan requirements. Sidelobes are less than -17 dB relative to the main beam peak.

Figure 22-12 shows an electromechanically scanned array called a VICTS (variable inclination continuous transverse stub) array, an innovative and practical solution for providing scan in two planes from a single horizontal array. The center photo of a Ku-band array shows a number of long slots cut through the surface of a circular plate. These slots are very short parallel plate waveguides that support only the lowest order (TEM)



(a)



(b)

FIGURE 22-11 EHF hybrid scan antenna with flush radome: (a) array mounting configuration and (b) EHF hybrid scan array (Courtesy of Hazeltine Corporation)

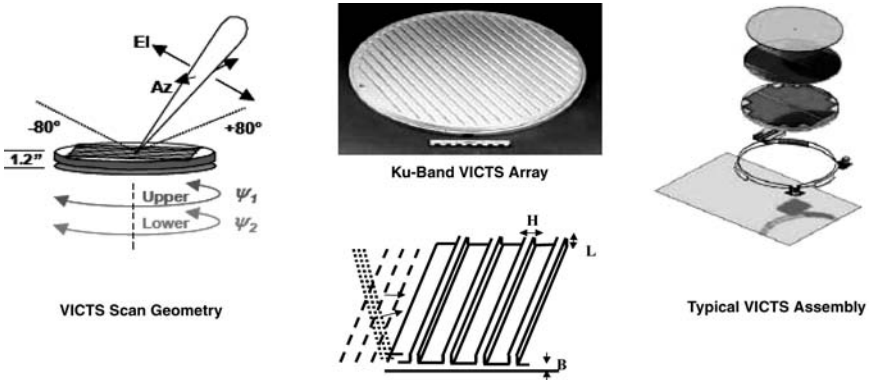


FIGURE 22-12 Mechanically scanned VICTS array (Courtesy of ThinKom Solutions, Inc., Torrance, CA)

waveguide mode. A second plate is indicated in the sketches to the right and left. This plate can rotate with or independently from the slotted plate. The sketch below the photo shows the upper plate as a series of radiating waveguide stubs of height L and width H being fed by a parallel plate waveguide of height B . This is the basic configuration of the continuous transverse stub array that has been the subject of several publications and patents.^{41,42} The dashed and dotted lines indicate different directions (and wavelengths) of the incident waves in the parallel plate region. The wavefronts indicated by dashed lines are parallel to the rows of stub radiators (slots), and so radiate broadside. The wavefronts shown by the dotted lines are at an angle to the stubs and excite radiation from the stubs that have a progressive time dependence to scan the stub radiation from broadside.

In the VICTS array this action is obtained by rotating the bottom plate relative to the top plate and can scan to 80° from the broadside. Rotating both plates together will then scan the radiated beam in the azimuth plane.

Fully electronically scanned arrays have also been developed for airborne satcom at SHF and EHF frequencies. Since in most cases the aircraft radii of curvature are large compared with the array size, the arrays can be planar and flush mounted without having to conform to the cylinder. This technology has been under development since the mid-1980s by the U.S. Air Force and NASA. These developments have led to a full line of arrays, including the three satcom receive arrays shown in Figure 22-13 at 8, 12, and 20 GHz.

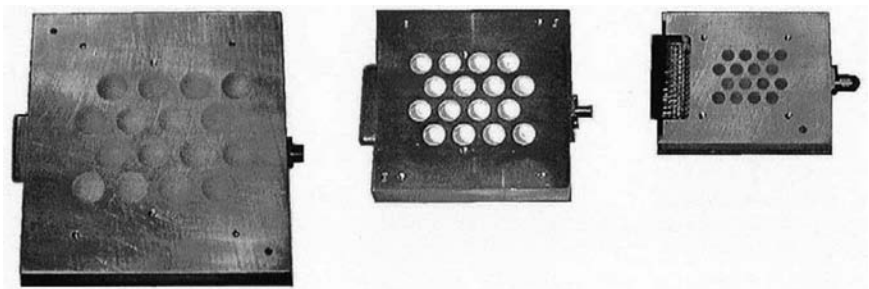


FIGURE 22-13 Low-profile arrays for satcom receive applications (Courtesy of Boeing Aircraft Corporation)

The small arrays shown here have only 16 elements and depth of 1.4 inches at 8 GHz, 0.94 at 12 GHz, and 0.75 inch at 20 GHz. The array elements are dielectric loaded circular waveguides excited by switchable left- and right-handed circularly polarized modules that have four phase shifter bits and an extra bit to select the polarization.

Elements and Arrays on Cylinders

When the array size approaches the vehicle radius of curvature, the behavior of the array element pattern becomes very different from either the isolated element pattern or the element pattern in a planar array. This property is well documented in the case of circular and cylindrical arrays, which have been the subject of more detailed studies than other conformal arrays. It has long been known that by using the symmetry properties of the array, you can avoid the solution of N simultaneous integral equations for an N -element circular array. Instead, you can superimpose the solution of N independent integral equations, one for each of the solutions with periodicity $2\pi/N$. This method has been used by Tillman⁴³ and King et al⁴⁴ for dipole and monopole arrays and by Borgiotti and Balzano¹³ for waveguide arrays for active elements covering a sector on an infinite cylinder. The results of these and other theoretical studies have revealed the existence of potentially severe ripples in cylindrical-array element patterns, as shown in Figure 22-14a. Sureau and Hessel¹⁴ have used asymptotic methods to show that the radiated pattern can be considered the superposition of a space-wave contribution and several creeping waves that propagate around the cylinder and radiate away from the cylinder in the forward direction. The space-wave and actual patterns are shown in the figure for radiation at two different angles θ measured from the cylinder axis. The ripple amplitude decreases with cylinder radius and increases with frequency. The work by Herper et al,¹⁶ indicated in Figure 22-14b, has shown that smoother element patterns can be obtained by limiting element spacing to $\lambda/2$. The improvement is attributed to the fact that the interference of the fast creeping-wave grating lobe¹⁶ with the direct ray decreases as the azimuth spacing is reduced and disappears at 0.5λ . This work showed that low-sidelobe circular arrays are feasible with proper concern for element design.

With active control, conformal cylindrical arrays can operate in omnidirectional or directional modes, achieving coverage over most of the sphere. However, to achieve efficient scanning it is necessary to include mutual coupling effects in element and grid design.

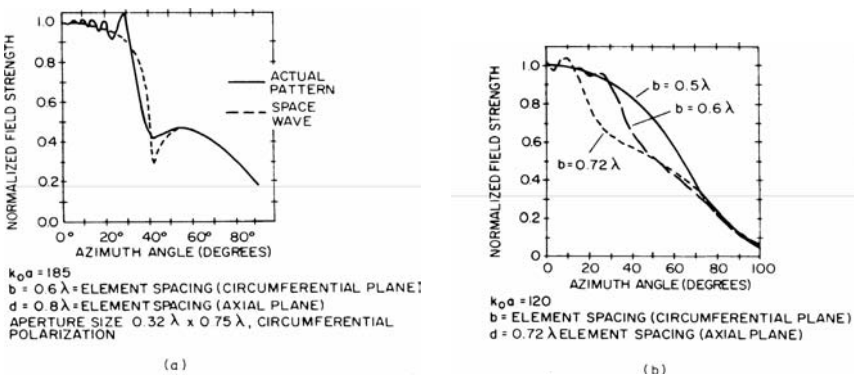


FIGURE 22-14 Circumferential element patterns: (a) waveguide array on cylinder (after Sureau and Hessel¹⁴ © Artech House 1972) and (b) dipole array in cylinder (after Herper et al¹⁶ © IEEE 1980)

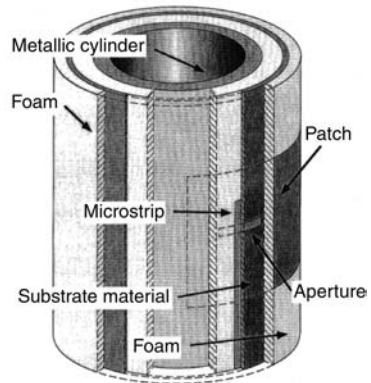
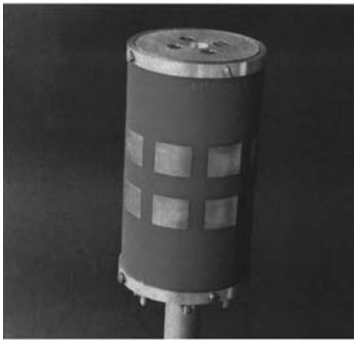


FIGURE 22-15 Cylindrical array of microstrip patches (after Löffler et al.⁴⁵ © IEEE 1999)

Figure 22-15⁴⁵ shows an example of a small dual function array of microstrip patches mounted on a circular cylinder. The modeling was accomplished using a spectral Green's function and the method of moments.

The array consists of eight vertical subarrays of two elements each and operates at C-band over more than 19 percent bandwidth. Each subarray has 10–11-dB gain over the operating bandwidth with a vertical half power beamwidth of 42°. Figure 22-15b shows a cross section of the cylindrical configuration with one element excited. The array consists of a metal cylinder 35 mm in diameter covered by an inner layer of foam 14 mm thick. This layer abuts the microstrip feedlines mounted on a 0.5-mm Duroid substrate. The outer side of the Duroid is metalized, with an aperture that couples the energy from the microstrip line through to the patch antennas. The patch itself has a 4-mm substrate layer of foam separating it from the aperture. The ability to model and build such a complex structure, designing to avoid resonances caused by modes that can propagate around the cylindrical structure, is testimony to the maturity of the analysis and software.

Feed Systems and Commutators for Cylindrical Arrays

Early experimental and theoretical studies concerned arrays that were built with dipole or waveguide² elements, with the array extending completely around the cylinder. Switching was accomplished by a variety of networks that were contained inside the cylinder, some of which are described in this section. Figure 22-16 shows a circular array of waveguide horns developed by the Naval Electronics Laboratory Center⁴⁶ and used in a study of fundamental radiating properties as well as a comparison of several kinds of switching networks. Other studies dealing with full cylindrical or ring arrays are described in the basic reference by Davies.¹²

When the array extends completely around the cylindrical surface, it is necessary to devise a feed network to rotate the illuminated spot continuously through multiples of 360°. As noted by Provencher,² Hill,⁴⁷ and Holley et al.,⁴⁸ a number of mechanical and electrical networks have been devised to accomplish that task. Mechanical commutators such as the waveguide commutator of Figure 22-17a are appropriate for systems with continuous scanning over multiple 360° increments. Such commutators are often seen as viable low-cost solutions for a number of applications, and they can provide scanned patterns relatively free

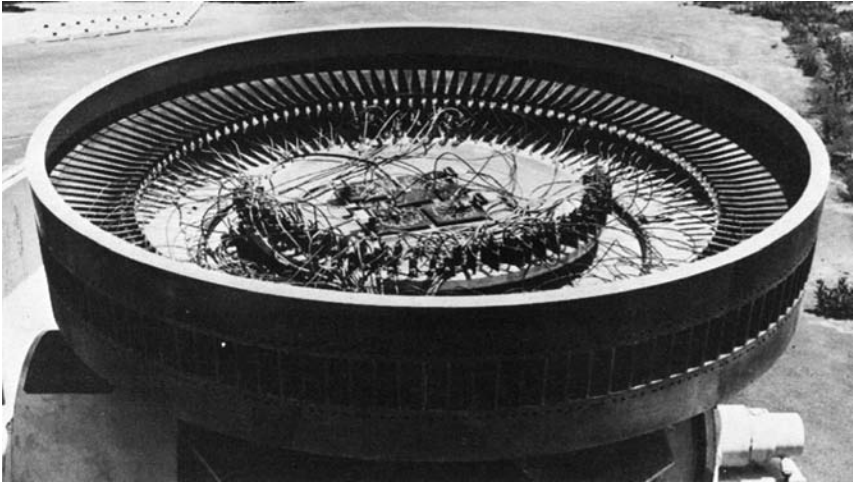


FIGURE 22-16 Ring array (after Boyns et al⁴⁶ © IEEE 1970)

of modulation effects if care is taken to assure that the rotating (rotor) feed elements have spacing different from those of the stationary (stator) pickup elements. This design procedure can assure that a minimum of elements is misaligned at any one time and so reduces radiation modulation. Power distribution within the rotor is achieved by using coaxial or waveguide corporate networks or radial power dividers.

Other investigators have considered switching networks for commuting an amplitude and phase distribution around a circular array. Giannini⁴⁹ describes a technique that uses a bank of switches to bring a given illumination taper to one sector of the array (usually a 90° to 120° arc) and a set of switches to provide fine beam steering between those characteristic positions determined by the sector-switching network. The circuit for a 32-element array is shown in Figure 22-17b; it requires eight phase shifters and 12 transfer switches (double-pole, double-throw) and achieves sector selection by using eight single-pole, four-throw switches. This network excites an eight-element quadrant of the array that can be moved in increments of one element to provide coarse beam steering. Fine steering is provided by the phase controls.

Several lens-fed circular arrays have been implemented by using *R*-2*R* Luneburg and geodesic lenses.⁴⁸ The *R*-2*R* lens (see Figure 22-17c) described by Boyns et al⁴⁶ forms as many beams as there are elements in the array. However, like all switched lens systems, it does not admit to fine steering (selecting angles with a separation less than the angular separation between radiating elements on the cylinder) unless additional phase controls are added to each element.

As pointed out by Holley et al,⁴⁸ lens systems can be adapted for fine steering by using an amplitude illumination with a moveable phase center. The net result of a progressive phase tilt at the input of the Butler matrix is to synthesize intermediate beams from a composite of the available lens beams and so to provide high-quality fine steering of the lens-radiated pattern.

There have been a number of developments in the area of multimode electronic commutators for circular arrays. These systems derive from techniques similar to that first used by Honey and Jones⁵⁰ for a direction-finding-antenna application in which several modes of a biconical antenna were combined to produce a directional pattern with full

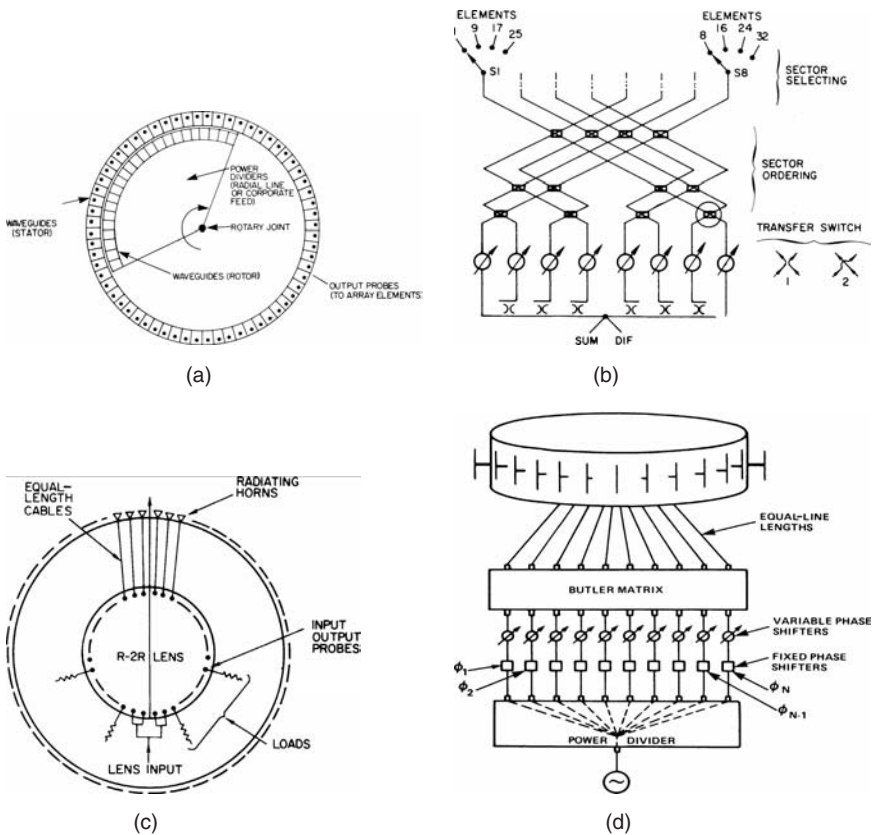


FIGURE 22-17 Circular-array feed systems: (a) commutator (waveguide), (b) switch network (after Giannini⁴⁹ © IEEE 1969), (c) R-2R lens (after Boyns et al.⁴⁶ © IEEE 1970), and (d) matrix scanning system (after Sheleg⁵⁴ © IEEE 1968)

360° azimuthal rotation. Bogner⁵¹ and Irzinski⁵² specifically address the use of such commutators, which must be combined with phase shifters and switches at each element of the array. The phase shifters provide collimating and fine steering, and the switches are used to truncate the illumination so that only a finite sector of the array is used at a time, a procedure that is required for sidelobe control.

Several more sophisticated types of electronic switches for circular arrays are based on a concept, originally proposed by Shelton⁵³ and developed by Sheleg,⁵⁴ that uses a matrix-fed circular array with fixed phase shifters to excite current modes around the array and variable phase shifters to provide continuous scanning of the radiated beam over 360°. The geometry is shown in Figure 22-17d. Skahil and White⁵⁵ later proposed a concept that excites only that part of the circular array that contributes to formation of the desired radiation pattern. The array is divided into a given number of equal sectors, with each sector excited by a Butler matrix and phase shifters. With either of these circuits, sidelobe levels can be lowered by weighting the input excitations to the Butler matrix. The technique of Skahil and White was demonstrated by using an 8 × 8 Butler matrix, eight phase shifters, and eight single-pole, four-throw switches to feed four 8-element sectors of a 32-element array. The design sidelobes were −24 dB, and measured data showed sidelobes below −22 dB.



FIGURE 22-18 Circular array (*Courtesy of Anaren Corporation*)

Rehnmark⁵⁶ showed that you can use the measured phase modes of the received signal to determine its direction of arrival with high accuracy. He observed that since the n th higher order phase mode angular phase progression around the cylindrical or circular array varies n times faster than that of the lowest order ($n = 1$) mode, then the phase progression difference between the $n = +2$ and $n = -2$ modes would be four times as fast as that of the lowest mode, and this rapid phase characteristic should allow four times the angle resolution. He postulated that the resulting ambiguity could then be resolved using the lowest order mode. The circular array of Figure 22-18 is a broadband direction finding antenna based on this concept.

As pointed out by Holley et al.,⁴⁸ lens systems can be adapted for fine steering by using an amplitude illumination with a movable phase center. The net result of a progressive phase tilt at the input of the Butler matrix is to synthesize intermediate beams from a composite of the available lens beams and so to provide high-quality fine steering of the lens-radiated pattern.

22-4 PATTERN SYNTHESIS FOR CYLINDRICAL SECTOR ARRAYS

Among the difficulties with circular sector array synthesis are that the elements of the array have element-pattern maxima at different angles according to their location and that the elements do not lie in a plane. This situation is depicted in Figure 22-19a and precludes use of all the standard synthesis methods. You can, however, control near sidelobes by projecting the array element locations and element patterns onto a plane tangent to the cylinder, as shown

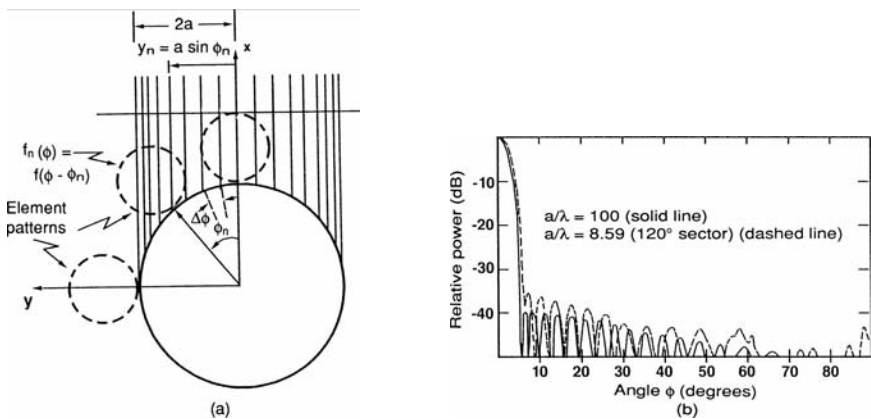


FIGURE 22-19 Synthesis of circular sector array patterns: (a) projection of array-element locations and (b) synthesized patterns with Taylor $n_{\text{bar}} = 8$, -40-dB sidelobe pattern, $a/\lambda = 100$ (solid), $a/\lambda = 8.59$ (dashed)

in Figure 22-19a. In this process you need to account for the nonuniform spacing that results from the projection, but the element patterns tend to account for this nonuniformity.

Figure 22-19a depicts a circular sector array with elements located equally spaced around the circumferential sector with angular separation $\Delta\phi$. Projected onto the array tangential plane, the element locations of an array are

$$\begin{aligned} y_n &= a \sin \phi_n \\ &= a \sin (n \Delta\phi) \end{aligned}$$

The projection tends to make the points y_n closer together near the ends of the array, but to first-order the element pattern tends to make up for that effect. For relatively large ϕ_n , the projected spacing between elements varies approximately like $\cos \phi_n$, and so the density of elements has a $1/\cos \phi_n$ dependence. If the array element patterns vary like $\cos \phi_n$, then to first-order the projected array weighting will be correct without altering the weights from those of a linear array with omnidirectional element patterns. If the element pattern is very different from a cosine, that fact must be included in the chosen weights.

Figure 22-19 shows several patterns synthesized by projecting a -40-dB Taylor pattern with $n_{\text{bar}} = 8$ onto cylinders with various radii and 36 elements. Element spacings were restricted to 0.5λ to avoid the grating-lobe sidelobes. The patterns relative to the local cylinder normal at ϕ_n in the ϕ plane have the form

$$f_n(\phi) = \cos(\phi - \phi_n)$$

The two curves show a nearly perfect Taylor pattern at radius $a = 100\lambda$ (solid) and a dashed pattern at $a = 8.59\lambda$ corresponding to an array that occupies a 120° sector of the cylinder. Significant changes in these patterns are beam broadening that results from shortening the array length, and an increased sidelobe level near the main beam, with the first sidelobe starting to merge with the main beam for small cylinders. Even for this small cylinder and the array wrapped over a 120° sector, the pattern is still not severely distorted.

A synthesis method first suggested by Sureau and Keeping⁵⁷ and further developed by Dufort⁵⁸ and Olen and Compton⁵⁹ is readily applicable to conformal arrays on generalized surfaces. The technique uses adaptive optimization algorithms to form the array pattern in the presence of closely spaced sources of interference that are tailored or iterate⁵⁹ to achieve the desired pattern. Figure 22-20 shows results from Sureau and Keeping⁵⁷ for a circular sector array of 32 identical elements disposed over a 120° sector of a cylinder and displaced 0.55λ between elements. The geometry shown in Figure 22-20a shows the 32 excited elements of a 96-element circular array. A total of 372 sources of interference were uniformly distributed outside the main-beam window to control sidelobes. The measured element pattern was used in the calculations. Figure 22-20b shows that increasing the window width produces a set of patterns with progressively lower sidelobes but decreasing aperture efficiency. Sureau and Keeping also investigated varying the interference weights to control sidelobe decay and the use of asymmetric weights for monopulse pattern control.

Other recent developments in synthesis include the use of nonlinear optimization⁶⁰ and least squares iterative procedures for optimizing over pattern parameters and frequency.⁶¹

Figure 22-21 shows results due to Bucci et al.⁶² Illustrating the synthesis of a \csc^2 pattern using an array of 21×15 elementary magnetic dipoles on a cylinder of radius 5λ . The 15 rows of elements are spaced $\lambda/2$ apart in the axial direction, and the 21 columns are spaced 0.7λ apart along the circumference of the cylinder so that an outer semicircle is occupied by the array. The technique employed in this synthesis is an iterative projection method, using a mask function to limit the upper and lower boundaries of the pattern function. In this method, the projection takes place sequentially between two vector spaces, one being the set of the squared amplitudes of all copolar and cross-polar components of all radiated patterns, and the other being the set of all patterns within the mask limits.

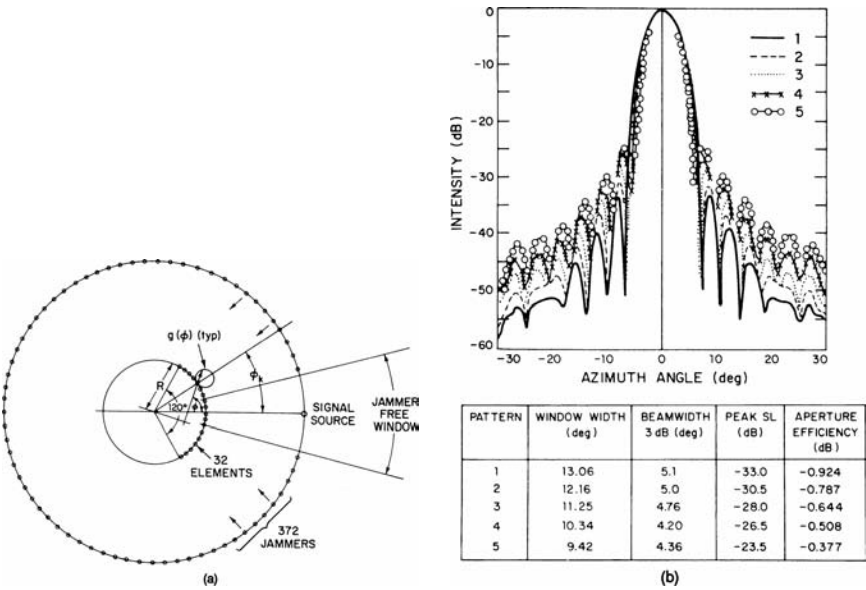


FIGURE 22-20 Synthesis of circular sector array patterns using the method of Sureau and Keeping⁵⁷: (a) geometry showing array, jammer locations, and jammer free window and (b) computed patterns and characteristics for various width jammer free windows (a and b after Sureau and Keeping⁵⁷ © IEEE 1982)

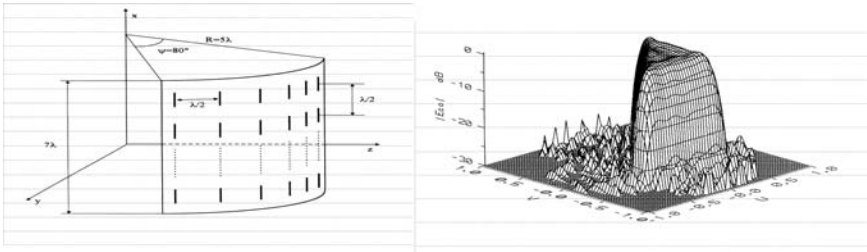


FIGURE 22-21 Synthesis of csc^2 patterns via projection (after Bucci et al⁶² © IEEE 1995)

This approach has been used successfully in a number of planar array studies and provides a convenient method with which to synthesize radiation patterns subject to useful constraints like the dynamic range of current amplitudes or reduced number of controls etc.

Figure 22-21a shows the elevation pattern shape and does not demonstrate the conformal synthesis features of the method since this plane is orthogonal to the curvature of the geometry. It does, however, illustrate the use of a mask to bound the synthesis. The procedure synthesized a full triangular shape (see Figure 22-21b) when projected onto direction cosine space. With all its generality, the procedure should find continued application to the synthesis of conformal arrays.

REFERENCES

1. W. H. Kummer, "Preface to Special Issue on Conformal Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-22, no. 1 (January 1974): 1–3.
2. J. H. Provencher, "A Survey of Circular Symmetric Arrays," *Phased Array Antennas*, A. Oliner and G. Knittel (eds.) (Dedham, MA: Artech House, Inc., 1972): 315–322.
3. P. Pathak and R. Kouyoumjian, "An Analysis of the Radiation from Apertures in Curved Surfaces by the Geometrical Theory of Diffraction," *IEEE Proc.*, vol. 62, no. 11 (November 1974): 1438–1447.
4. A. Hessel, "Mutual Coupling Effects in Circular Arrays on Cylindrical Surfaces—Aperture Design Implications and Analysis," *Phased Array Antennas*, A. Oliner and G. Knittel (eds.) (Dedham, MA: Artech House, Inc., 1972): 273–291.
5. L. Schwartzman and J. Stangel, "The Dome Antenna," *Microwave J.* (October 1974): 31–34.
6. H. Steyskal, A. Hessel, and J. Shmoys, "On the Gain-Versus-Scan Trade-Offs and the Phase Gradient Synthesis for a Cylindrical Dome Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-27 (November 1979): 825–831.
7. L. Josefsson and P. Persson, *Conformal Array Antenna Theory and Design* (Hoboken, NJ: IEEE Press, Wiley Interscience, 2006).
8. M. Kanno, T. Hashimura, and T. Katada, "Digital Beam Forming for Conformal Active Array Antenna," *1996 IEEE International Symposium on Phased Array Systems and Technology Digest* (October 15–18, 1996): 37–40.
9. M. Kanno, T. Hashimura, T. Katada, T. Watanabe, S. Iio, H. Soga, T. Nakada, and N. Miyano, "An Active Conformal Array Antenna with Very Small Thickness," *1996 IEEE International Symposium on Phased Array Systems and Technology Digest* (October 15–18, 1996): 155–158.
10. E. Rai, S. Nishimoto, T. Katada, and H. Watanabe, "Historical Overview of Phased Array Antenna for Defense Application in Japan," *1996 IEEE International Symposium on Phased Array Systems and Technology Digest* (October 15–18, 1996): 217–221.

11. H. Steyskal, "Circular Array with Frequency Independent Pattern," *AP-S International Symp.*, (1989): 1477–1480.
12. D. E. N. Davies, "Circular Arrays," The Handbook of Antennna Design, Chap. 12, vol. II, A. W. Rudge, K. Milne, A. D. Olver, and P. Knight (eds.), *IEE EM Waves Series* (London: Peter Perigrinus Ltd., 1983).
13. G. V. Borgiotti and Q. Balzano, "Analysis of Element Pattern Design of Periodic Array of Circular Apertures on Conducting Cylinders," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (September 1972): 547–553.
14. J. C. Sureau and A. Hessel, "Realized Gain Function for a Cylindrical Array of Open-Ended Waveguides," *Phased Array Antennas* A. Oliner and G. Knittel (eds.) (Dedham, MA: Artech House, Inc., 1972), pp. 315–322.
15. Q. Balzano and T. B. Dowling, "Mutual Coupling Analysis of Arrays of Aperture on Cones," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (January 1974): 92–97.
16. J. C. Herper, C. Mandarino, A. Hessel, and B. Tomasic, "Performance of a Dipole Element in a Cylindrical Array—A Modal Approach," *IEEE AP-S Inst. Symp.* (June 1980): 162–165.
17. T. M. Habashy, S. M. Ali, and J. A. Kong, "Input Impedance and Radiation Pattern of Cylindrical-Rectangular and Wraparound Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, vol. 38 (May 1990): 722–731.
18. K. Luk, K. Lee, and J. Dahele, "Analysis of the Cylindrical-Rectangular Patch Antenna," *IEEE Trans Antennas Propagat.*, vol. AP-37 (February 1989): 143–147.
19. R. Rostan, G. Gottwald, and W. Wiesbeck, "Design and Performance of Conformal Microstrip Patch Arrays on Cylindrical Surfaces," Proc of the 24th European Microwave Conference EuMC '94, Cannes, France, September 5–8, 1994: 1756–1761.
20. J. R. Descardecı and A. J. Giarola, "Microstrip Antenna on a Conical Surface," *IEEE Trans. Antennas and Propagat.*, vol. 40, no. 4 (April 1992): 460–468.
21. J.-M. Jin, J. A. Berrie, R. Kipp, and S.-W. Lee, "Calculation of Radiation Patterns of Microstrip Antennas on Cylindrical Bodies of Arbitrary Cross Section," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 1 (January 1997): 126–132.
22. P.-S. Kildal and J. Sanford, "Analysis of Conformal Antennas by Using Spectral Domain Techniques for Curved Structures," COST 245/ESTEC Workshop on Active Antennas, June 1996.
23. Z. Sipus, M. Lanne, and L. Josefsson, "Moment Method Analysis of Circular Cylindrical Array of Waveguide Elements Covered with a Multiplayer Radome," *Microwaves, Antennas and Propagat.*, IEE Proceedings, vol.153, no.1 (February 6, 2006): 29–37.
24. J. M. Song, C. C. Lu, W. C. Chu, and S. W. Lee, "Fast Illinois Solver Code (FISC)" *IEEE Antennas and Propagation Magazine*, vol. 40, no. 3 (June 1998): 27–31.
25. W. C. Chew, J. M. Jin, E. Michielssen, and J. M. Song (eds.), *Fast and Efficient Algorithms in Computational Electromagnetics* (Norwood, MA: Artech House, Inc., 2001).
26. T. G. Jurgens, A. Taflove, K. Umashankar, and T. G. Moore, "Finite-Difference Time-Domain Modeling of Curved Surfaces," *IEEE Trans. Antennas Propagat.*, vol. 40, no. 4 (April 1992): 357–365.
27. T. Kashiwa, T. Onishi, and I. Fukai, "Analysis of Microstrip Antennas on a Curved Surface Using the Conformal Grids FD-TD Method," *IEEE Trans. Antennas Propagat.*, vol. 42, no. 3 (March 1994): 423–432.
28. J. L. Volakis, A. Chatterjee, and L. C. Kempel, *Finite Element Method for Electromagnetics*, (London: IEEE Press/Oxford University Press, 1988).
29. T. Özdemir and J. L. Volakis, "Triangular Prisms for Edge-Based Vector Finite Element Analysis of Conformal Antennas," *IEEE Trans. Antennas Propagat.*, vol. 45, no 5 (May 1997): 788–797.
30. T. Özdemir, M. W. Nurnberger, J. L. Volakis, R. Kipp, and J. Berrie, "A Hybridization of Finite-Element and High-Frequency Methods for Pattern Predication for Antennas on Aircraft Structures," *IEEE Antennas and Propagation Magazine*, vol. 38, no. 3 (June 1996): 28–38.

31. K. E. Golden et al, "Approximation Techniques for the Mutual Admittance of Slot Antennas in Metallic Cones," *IEEE Trans Antennas Propagat.*, vol. AP-22 (January 1974): 44–48.
32. H. Steyskal, "Analysis of Circular Waveguide Arrays on Cylinders," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (1977): 610–616.
33. P. Munk and P. H. Pathak, "A UTD Analysis of the Radiation and Mutual Coupling Associated with Antennas on a Smooth Perfectly Conducting Arbitrary Convex Surface with a Uniform Material Coating," *IEEE APS International Symposium*, vol. 1 (1996): 696–699.
34. C. Demirdag and F. G. Rojas, "Mutual Coupling Calculations on a Dielectric Coated PEC Cylinder Using a UTD-Based Green's Function," *IEEE APS International Symposium 1997 Digest* (July 1997): 1525–1527.
35. P. Persson and L. Josefsson, "Calculating the Mutual Coupling Between Apertures on Convex Cylinders Using a Hybrid UTD-MOM Method," *IEEE/APS International Symposium*, July 1999.
36. C. H. Walter, *Traveling Wave Antennas* (New York: McGraw-Hill, 1965), pp. 121–122.
37. L. B. Corey, "A Graphical Technique for Determining Optimal Array Antenna Geometry," *IEEE AP-S Trans.*, vol. AP-33, no.7 (July 1985): 719–726.
38. R. J. Mailloux and W. G. Mavroides, "Hemispherical Coverage of Four-Faced Aircraft Arrays," *RADC-TR-79-176*, May 1979.
39. R. G. Kouyoumjian and P. H. Pathak, "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," *IEEE Proc.*, vol. 62 (November 1974): 1448–1461.
40. J. F. Pederson, "Measured Antenna Performance of an EHF Hybrid Scan Array for Milsatcom Applications," *Proceedings of the 1988 Antenna Applications Symposium*, *RADC-TR-879-121*, vol. 2: 461–483.
41. W. W. Milroy, "Advanced Broadband Access Applications of the Continuous Transverse Stub (CTS) Array," *International Conference on Electromagnetics in Advanced Applications (ICEAA01)*, Torino, Italy, September 10–14, 2001.
42. W. W. Milroy, S. B. Coppedge, and A. C. Lemons, Variable Inclination Continuous Transverse Stub Array (July 19, 2005): U. S. Pat. # 6,919,854 B2.
43. J. T. Tillman, Jr., "The Theory of Circular Array Antennas," *Eng. Experiment Station Rep.* (Knoxville: University of Tennessee, 1968).
44. R. W. P. King, R. Mack, and S. S. Sandler, *Arrays of Cylindrical Dipoles*, Chap. 4 (London: Cambridge University Press, 1968).
45. D. Löffler, W. Wiesbeck, and B. Johansson, "Conformal Aperture Coupled Microstrip Phased Array on a Cylindrical Surface," *IEEE Int. Symposium*, July 11–16, 1999.
46. J. E. Boyns, C. W. Gorham, A. D. Munger, J. H. Provencher, J. Reindel, and B. I. Small, "Step-Scanned Circular Array Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-18, no. 5 (September 1970): 590–595.
47. R. T. Hill, "Phased Array Feed Systems: A Survey," *Phased Array Antennas*, A. Oliner and G. Knittel (eds.) (Dedham, MA: Artech House, Inc., 1972), pp. 197–211.
48. E. Holley, E. C. DuFort, and R. A. Dell-Imagine, "An Electronically Scanned Beacon Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-22, no. 1 (January 1974): 3–12.
49. R. S. Giannini, "An Electronically Scanned Cylindrical Array Based on a Switching and Phasing Technique," *IEEE Antennas Propagat. Int. Symp. Dig.* (December 1969): 199–207.
50. R. C. Honey and E. M. T. Jones, "A Versatile Multiport Biconical Antenna," *IRE Proc.*, vol. 45 (October 1957): 1374–1383.
51. B. F. Bogner, "Circularly Symmetric R. F. Commutator for Cylindrical Phased Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-22, no. 1 (January 1974): 78–81.
52. E. P. Irzinski, "A Coaxial Waveguide Commutator Feed for a Scanning Circular Phased Array," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29, no. 3 (March 1981): 266–270.

53. P. Shelton, "Application of Hybrid Matrices to Various Multimode and Multibeam Antenna Systems," IEEE Washington Chap. PGAP Meeting, March 1965.
54. B. Sheleg, "A Matrix-Fed Circular Array for Continuous Scanning," *IEEE Proc.*, vol. 56, no. 11 (November 1968): 2016–2027.
55. G. Skahil and W. D. White, "A New Technique for Feeding a Cylindrical Array," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (March 1975): 253–256.
56. S. Rehnmark, "Instantaneous Bearing Discriminator with Omnidirectional Coverage and High Accuracy," *IEEE MTT-S International Symposium* (May 1980): 120–122.
57. J. C. Sureau and K. J. Keeping, "Sidelobe Control in Cylindrical Arrays," *IEEE Trans. Antennas Propagat.*, vol. 30, no. 5 (September 1982): 1027–1031.
58. E. C. Dufort, "Pattern Synthesis Based on Adaptive Array Theory," *IEEE Trans. Antennas Propagat.*, vol. AP-37, no. 8 (August 1989): 1011–1018.
59. C. A. Olen and R. T. Compton, "A Numerical Pattern Synthesis Algorithm for Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-38, no.10 (October 1990): 1666–1676.
60. Y.-C. Jiao, W.-Y. Wei, L.-W. Huang, and H.-S. Wu, "A New Low-Side-Lobe Pattern Synthesis Technique for Conformal Arrays," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 6 (June 1993): 824–831.
61. L. I. Valkelainen, "Iterative Least-Squares Synthesis Method for Conformal Array Antennas with Optimized Polarization and Frequency Properties," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 7 (July 1997): 1179–1185.
62. O. M. Bucci, G. D'elia, and G. Romito, "A Generalized Projection Technique for the Synthesis of Conformal Arrays," *IEEE –AP-S International Symp.*, (1995): 1986–1989.

Chapter 23

Millimeter-Wave and Terahertz Antennas

Rudy Emrick

Motorola

John L. Volakis

The Ohio State University

CONTENTS

23.1	MILLIMETER-WAVE AND TERAHERTZ APPLICATIONS	23-2
23.2	MILLIMETER-WAVE ANTENNAS	23-5
22.3	SUBMILLIMETER-WAVE AND TERAHERTZ ANTENNAS	23-12

23.1 MILLIMETER-WAVE AND TERAHERTZ APPLICATIONS

There are a number of definitions describing antennas operating in the millimeter-wave (MMW) or terahertz (THz) bands. Most commonly, antennas operating in frequencies whose wavelengths are measured in millimeters (30 GHz–300 GHz) are referred to as *millimeter-wave antennas*. On the other hand, the band between 300 GHz and 3 THz is referred to as the *submillimeter-wave band*, in which the corresponding wavelengths are measured in units less than one millimeter. One can say that a fair amount of freedom has been used in the literature when defining the MMW and THz bands. Some describe a device operating at 24 GHz as a millimeter-wave sensor and those operating around 300 GHz as THz devices. A possible cause for these inconsistencies may be the specialized use of MMW and THz devices. Historically, devices operating in the millimeter-wave, submillimeter-wave, and THz regimes have been limited to specialized applications, likely due to their high cost of realization. Specifically, at these higher frequencies there are a number of challenges, including (1) the availability of sources operating in these bands, (2) fabrication challenges of the small features required for these devices, (3) maintaining tolerance for the small features to achieve repeatable designs, and (4) fabrication costs.

It is certainly well-known that antenna dimensions scale inversely proportional to their frequency. However, at some point, the feature's size becomes a fabrication challenge, and traditional printed circuit-board techniques may no longer be applicable. Nevertheless, recent and ongoing advances in a number of fabrication technologies indicate that low-cost solutions at the MMW and THz frequencies should be possible.¹ Specifically, silicon transceivers can now support frequencies of 60 GHz and higher. That is, with the capability of traditional silicon devices continuing to trend higher,² the challenges of obtaining sources and transceiver ICs is effectively being addressed. To some extent, this moves the challenge in realizing MMW and THz devices to other components, including antennas. This chapter is, of course, specifically focused on the antenna design and fabrication for MMW and THz devices.

Applications for millimeter-wave and THz frequencies include wireless, radar, and imaging (see Figure 23-1). For antennas, it is often found that the design is application-specific where integration, loss, and gain requirements may vary among applications. An issue for the designer is also the allowed frequency band of operation. Specifically, the unlicensed spectrum at 60 GHz overlaps an oxygen absorption band that adds an additional 15 dB/km of attenuation. This is certainly a challenge for long communication ranges,

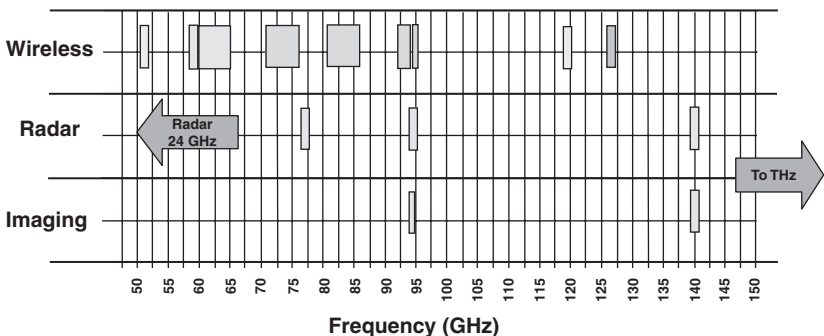


FIGURE 23-1 Millimeter-wave (MMW) spectrum and applications

but may not be of concern for shorter range applications, as is the case for wireless personal area networks. In fact, this atmospheric attenuation is even desired since it allows for frequency reuse and security of the band by limiting its reach to local areas of operations. From Figure 23-1, imaging applications are typically targeted for bands where there is a local minimum in atmospheric attenuation, as is the case with the 94 GHz and 140 GHz frequency bands. This is important, especially for passive imaging, as the desired signals can be extremely low in power and additional atmospheric attenuation can severely impact their reception. Next we consider the three applications areas separately.

Wireless

As wireless devices become more prevalent and the desire for increased data rates continues, there are challenges in the employed lower frequency spectrum (from 824 MHz for the cellular bands up to 2.4 GHz and 5 GHz for the ISM and 802.11 bands, see Figure 2-4 in Chapter 2). A solution is to use the large amounts of spectrum in the millimeter-wave frequencies. Wireless applications in these bands include wireless personal area networks (WPAN), point to point for backhaul applications, and the extension of fiber and hybrid fiber-coax systems beyond their current reach. The latter presents an advantage over wired networks as it can be cost prohibitive to extend them beyond their current coverage. Specifically, a large portion of unlicensed spectrum is available worldwide around 60 GHz for either WPAN or other applications such as backhaul. In the United States, easily and cheaply licensed spectrum is now available in two 5 GHz blocks near 75 and 85 GHz. It is, however, believed that spectrum targeted for commercial backhaul applications must be licensed so the operator has some recourse in case of interference.

Challenges faced in antenna development for wireless applications may differ significantly. For backhaul applications, the antennas are typically directional and physically large for acceptable performance over 1–2 km in range under typical rain-fade conditions. However, in the case of WPAN, the challenges range from low cost to sufficient gain for acceptable system-level performance allowing for high data-rate connectivity. To achieve multigigabit data rates at reasonable ranges, especially for non-line-of-sight (NLOS) conditions, it has been recently shown³ (using simple channel models) that antenna gains on the order of 20 dBi are likely required. Though these antennas will be smaller, an array may be required to achieve such large gains while accounting for expected chipset efficiencies. Figure 23-2 shows signal-to-noise (S/N) ratio as a function of separation distance under different propagation conditions and transmission systems such as single-input single-output (SISO) and multiple-input multiple-output (MIMO). The propagation channels included in Figure 23-2 include the usual line-of-sight (LOS) with multipaths and the non-line-of-sight (NLOS), referred to as LOS-blocked in the figure. An observation for Figure 23-2 is that MIMO (see Chapter 57) has a clear advantage in achieving higher data rates.

Radars

Millimeter-wave radar sensors are in early deployment for applications such as automotive anticollision systems. Virtually every major automaker now has an optional adaptive cruise control system that may become part of a future automotive safety system for collision warning/avoidance, lane departure warning, blind spot detection, backup, and parking aids. Such systems are already being used for parking assistance. Though there is debate on the future pervasiveness of anticollision systems, it is widely expected that these technologies will make driving safer. For example, without active avoidance, perimeter sensors may detect a situation having a high probability of collision and initiate pre-crash safety steps,

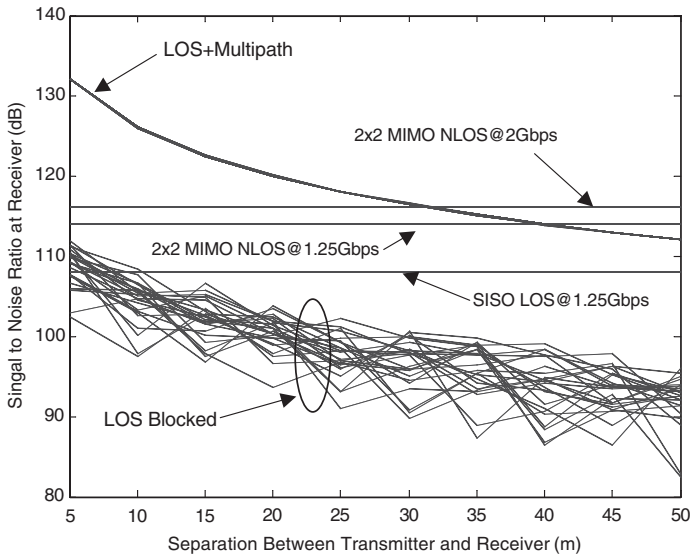


FIGURE 23-2 Signal-to-noise (S/N) ratio at the receiver as a function of separation distance between the transmitter and receiver. An antenna with 20-dBi gain was assumed (the horizontal lines show the required S/N ratios for the indicated data rates and configuration) (after R. M. Emrick and J. L. Volakis³ © IEEE 2006).

including pretensioning of seat belts, adjusting brake line pressure to shorten stopping distance, and so on. As displayed in Figure 23-3, there are a wide range of sensors likely to be used; these include millimeter-wave, optical, and ultrasonic sensors. Sensors operating in the 77-GHz range with a combination of 24-GHz sensors for shorter ranges (see Figure 23-1) are a likely option. Specifically, millimeter-wave frequencies are considered a good choice because they penetrate fog and heavy dust in addition to allowing for high resolution and being small enough so as not to affect vehicle appearance.

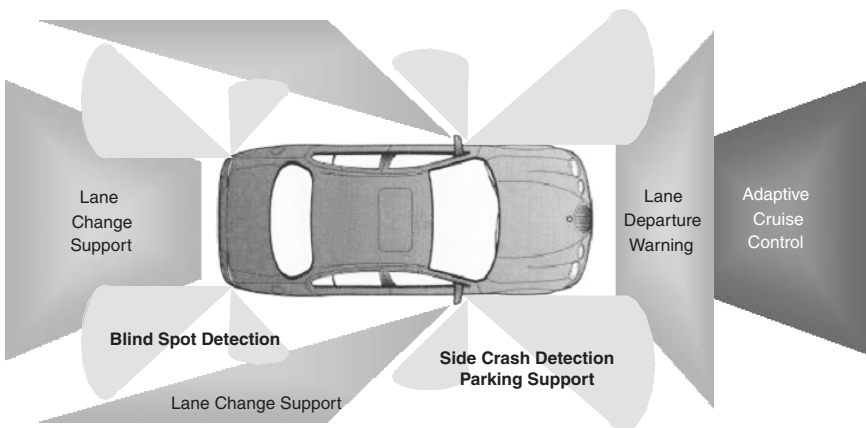


FIGURE 23-3 Automotive sensors for advanced safety systems

Imaging

The area of millimeter-wave and THz imaging has seen significant growth in a number of applications. A popular application is security screening for detection of concealed objects such as ceramic knives hidden beneath clothing⁴ (done passively). Another application is the use of passive millimeter-wave imagers for aircraft systems used by pilots to see landing areas under low visibility conditions (fog or sand storms). Such imaging sensors can operate at higher frequencies that extend to THz and may enable a wide range of systems for detection and diagnosis. A great potential for imaging using MMW and THz exists in the medical field. Applications include tumor recognition, disease diagnosis, recognition of protein structural states, detection/imaging of tooth decay, and so on.⁵

23.2 MILLIMETER-WAVE ANTENNAS

As mentioned previously, challenges associated with antennas at high frequencies include fabrication tolerances, efficiency/losses, and cost, though these are all somewhat inter-related. As is often the case, it can be very difficult to optimize each of these parameters simultaneously, and a number of approaches have already been pursued and developed. In this section, we discuss technologies used for the fabrication of different MMW antennas. We discuss these under the categories of (1) waveguide antennas, (2) printed planar antennas, and (3) on-chip antennas, which may be fabricated using micromachining or semiconductor lithography techniques. Here, the first two approaches will be discussed with the category of on-chip antennas being discussed in a later section.

Waveguide Antennas

An approach to minimizing losses in millimeter-wave antennas is to use waveguide antennas. It is well-known that waveguides and waveguide antennas can provide higher performance and lower losses, but they do not typically align well with compact or low-cost solutions. Nevertheless, in this context, one can consider waveguide feeds and horns or other waveguide antennas, but hybrid approaches have also been explored. When fabricating such components for MMW applications, a lower-cost approach is to utilize metalized plastics,⁶ but many question the viability of this approach. Some challenges associated with the metalized plastics include the type of metallization, interfacing with other components, and maintaining electrical continuity. Also, the overall cost of this approach and other measures relating to integration bring additional challenges. For example, the design must insure that under temperature cycling the joining or movement between dissimilar materials does not degrade metallization and metal contacts.

In addition to using metalized plastics to reduce cost, casting, or electroforming, the antenna followed by gold or other plating is a viable alternative. Two millimeter-wave horn antennas are shown in Figure 23-4. The horn on the left is fabricated using precision assembly and brazing/welding techniques, and the horn on the right is fabricated by plating a cast



FIGURE 23-4 Horn antennas, connected to a WR15 waveguide, operating at millimeter-wave frequencies. The horn antenna on the left is precision assembled whereas the horn on the right is cast and plated for lower cost.

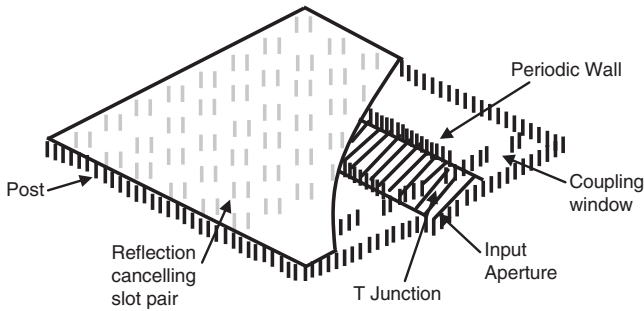


FIGURE 23-5 Parallel plate slot array to improve manufacturability. Dielectric constant of the employed material is 2.17 and an efficiency of 29% was achieved (after J. Hirokawa and M. Ando⁷ © IEEE 1998).

base metal. The electrical performance of the two horns shown in Figure 23-4 is very similar (within one to two tenths of a dB). However, casting (with plating) can significantly reduce fabrication cost from one to several orders of magnitude, depending on the fabrication volume.

An example approach to fabricating waveguide antennas for volume manufacturing is to employ parallel plate slot arrays,⁷ as shown in Figure 23-5. In this case, losses may be greater since a dielectric must be used for filling the waveguide. Of course, lower dielectric constants are preferred to avoid further reduction in the metallization dimensions, thus reducing fabrication costs and ohmic losses (issues related to high dielectric constants or thin substrates will be discussed later in more detail).

An additional example of a waveguide feed is shown in Figure 23-6. In this example,⁸ a waveguide is used to feed the microstrip antenna, and this may be a good compromise for certain applications since losses in the MMW feeding networks can be significant. The feed shown in Figure 23-6 is rather standard and employs an aperture in the broadside of the rectangular waveguide positioned under a microstrip antenna. Challenges for an approach like this, as in other cases, are likely to increase as the frequency of operation continues to increase since at higher frequencies, the dimensions of the aperture can become quite small, and placement of the microstrip antenna properly over the aperture can become a dominant challenge.

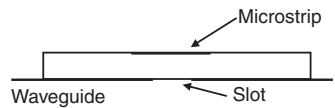


FIGURE 23-6 Waveguide to microstrip antenna coupling using an aperture at the broadside to feed the microstrip antenna (after D. Pozar⁸ © IEEE 1996)

Printed Planar Antennas

As may be expected, there is strong interest in printed planar antennas for millimeter-wave applications. At lower frequencies, these are very easy to fabricate and are compatible with other portions of a transceiver that may be implemented on the same substrate allowing for greater integration. There are a range of substrates used for planar antennas, but there are some that appear to have garnered the greatest interest based on the volume of current work and technical publications. The most commonly used substrates are currently low temperature cofired ceramic (LTCC), liquid crystal polymer (LCP), and silicon. Interest in LTCC and LCP is fairly wide because their electrical properties are some of the best at millimeter-wave frequencies and are also well-suited for integration with other portions of a transceiver. Silicon is certainly of high interest for on-chip antenna integration.

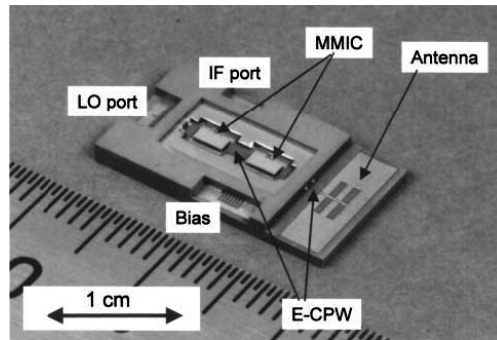


FIGURE 23-7 Double slot antenna implemented on an LTCC package (after K. Maruhashi et al⁹ © IEEE 2000)

Packages with integrated antennas using LTCC have already been demonstrated, and there are a number of options available in the type and implementation of printed planar antennas utilizing LTCC. A double slot antenna⁹ on LTCC is shown in Figure 23-7. In this case, the antenna is formed using several layers to produce a multilayer package delivering a gain of 4 dBi for the double slot antenna. This example represents good feasibility for antenna integration with the transceiver package. However, having the antenna and transceiver close to or colocated with the transceiver is not always preferred. Depending on the application and the transceiver, it may be preferable to have the antenna located where it can provide the best performance with the rest of the transceiver, for instance, located near its interfaces with the rest of the system. Having the antennas separated from the transceiver is, of course, common at lower frequencies, but at millimeter-wave frequencies interconnect losses between the transceiver and antenna are of concern. In this case, separations of many wavelengths can be expected for a package of a few millimeters in size. Using more available layers for antenna design with LTCCs is also an option, as demonstrated with the multilayer parasitic microstrip antenna array in Seki et al¹⁰ (see Figure 23-8). This antenna was formed using three LTCC layers and incorporates the feed element and two parasitic element layers. It delivered a measured gain of 7.17 dBi.

In addition to LTCCs, a substrate material that has been gaining interest is the liquid crystal polymer (LCP),¹¹ though LCP has been under consideration for much less time than LTCC. LCP has captured interest due to its good electrical characteristics. They are also considered very stable since LCP does not absorb moisture and is therefore near hermetic.

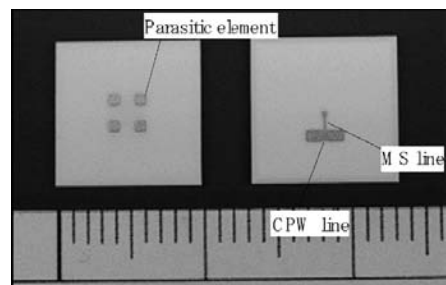
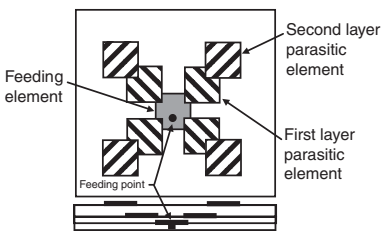


FIGURE 23-8 A multilayer parasitic microstrip antenna array implemented using LTCC (after T. Seki et al¹⁰ © IEEE 2005). Measured absolute gain for this antenna is 7.17 dBi.

TABLE 4-1 Comparison of Substrate Properties for LTCC, FR4, and LCP

	LTCC	FR4	LCP
Relative permittivity	5.6 (20 GHz)	4.5 (1 MHz)	2.9 (20 GHz)
Loss tangent	0.0012	0.02	0.002
Relative cost	Medium	Very low	Low

A comparison of LTCC, LCP, and FR4 is shown in Table 23-1. As shown, LCP has a loss tangent similar to LTCC with a lower dielectric constant. The latter may be an advantage at MMW frequencies, but as compared to LTCCs, it offers fewer dielectric constant choices.

Using LCP substrates with multiple layers, as may be done with LTCCs, a number of antenna possibilities exist. One such possibility is shown in Figure 23-9 where multiple LCP layers have been used to implement a vertical array of spiral antennas to achieve more than 12-dBi gain occupying a top surface area of only 1.2 mm. Additional increases in gain may be achieved by increasing the number of layers and increasing the number of elements. The operation of an array like this is similar to a Yagi-Uda antenna where the ground plane acts as the reflector element and the element above is the driver. Additional spiral elements above the driven element act as director elements. A challenge in this approach is feeding the driver element.

On-Chip Antennas

Since antennas are physically small at high frequencies, overcoming challenges with interconnect losses has forced implementations toward on-chip integration. A number of approaches beyond those that are part of standard packaging techniques (to minimize dielectric and metal losses) have been explored.¹²⁻¹⁵ In order to minimize losses, methods are often used to maximize the portion of the fields that are in air or to maximize the metallization used to print the transmission lines and antennas. However, since the characteristic impedance of transmission line is $Z_0 \approx (L/C)^{0.5}$, and the capacitance increases with the dielectric constant, as the capacitance increases the transmission line impedance is also lowered. To compensate for the increased capacitance, the width of the microstrip line can be decreased. This does raise the line impedance but also

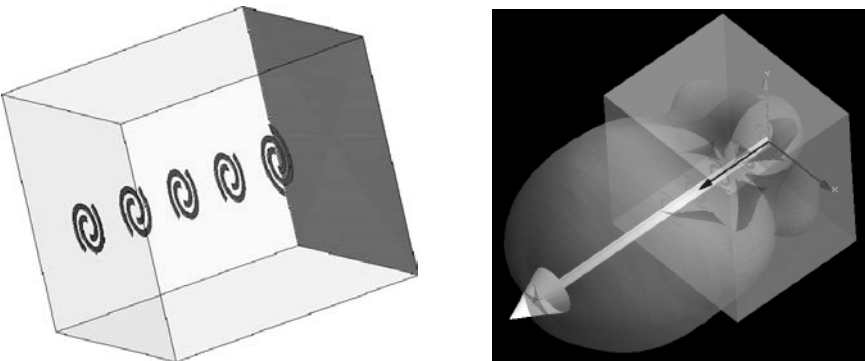


FIGURE 23-9 High-gain antenna utilizing multiple LCP layers to form a vertical array of spirals over a ground plane. The achieved gain is 12.3 dBi with a top surface occupying 1.2 mm, and computations were carried out using the Remcom XFDTD (after R. M. Emrick and J. L. Volakis³ © IEEE 2006).

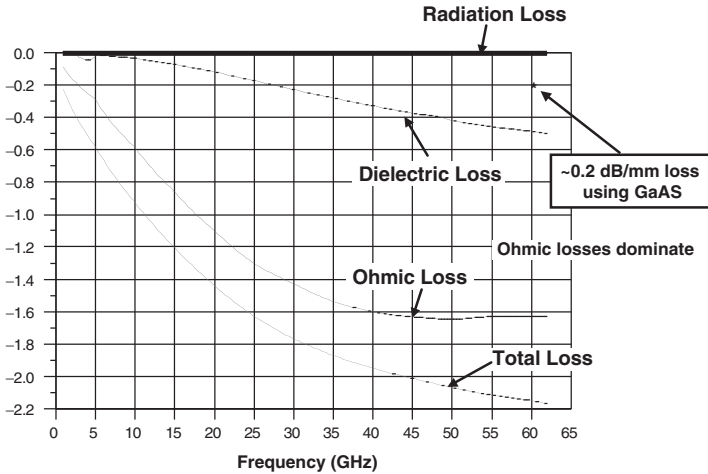


FIGURE 23-10 Transmission line losses in dB/mm using thin dielectrics. The SiO₂ substrate thickness $0.3\ \mu\text{m}$ is representative when implementing $50\ \Omega$ transmission lines in the top layers of silicon wafer processes. Losses for thicker GaAs substrates are also shown for comparison (use of higher dielectric constants produce similar effects).

increases ohmic losses associated with the transmission line. Figure 23-10 shows loss contributions due to thin dielectrics for a microstrip transmission line using simulations. The figure highlights the effects of ohmic losses and how these can become dominant as frequencies are increased (assuming a substrate with reasonable loss characteristics). The assumed relatively thin SiO₂ layer of $0.3\ \mu\text{m}$ is consistent with the upper layers available for the transmission line as part of a silicon process. For reference, losses of $\sim 0.2\ \text{dB/mm}$ are expected when using gallium arsenide (GaAs) with a commonly used substrate thickness of $100\ \mu\text{m}$.

An approach for reducing losses in structures and components by reducing the substrate dielectric used¹² is shown in Figure 23-11. By creating an air cavity below the structures, both the dielectric and ohmic losses can be reduced. Air cavities can be created to form such structures either by etching or by micromachining away materials from a thick substrate. Alternatively, the desired devices and components can be placed on a thin membrane either before or after being applied to a supporting structure. It is not clear that approaches like this will lend themselves to low-cost, high-volume manufacturing, though they are clearly of interest for specialized lower-volume applications. For such approaches to be viable for high volume, they should also be compatible with standard semiconductor processing. Additional issues may relate to the device packaging for long-term reliability as moisture or other contaminants that enter the air cavity can significantly affect performance. A larger scale example of this approach is shown in Figure 23-12. Figure 23-12 shows a two-dimensional array fabricated so the antenna element is suspended on a membrane over a cavity shaped as a horn. This approach allows for reduction in both dielectric and ohmic losses.

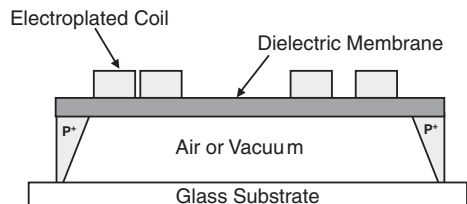


FIGURE 23-11 Approach for reducing losses by increasing the air volume near the structure or components of interest (after C. Nguyen et al¹² © IEEE 1998). This may have the effect of reducing dielectric losses in addition to possibly reducing ohmic losses due to decreased distributed capacitance.

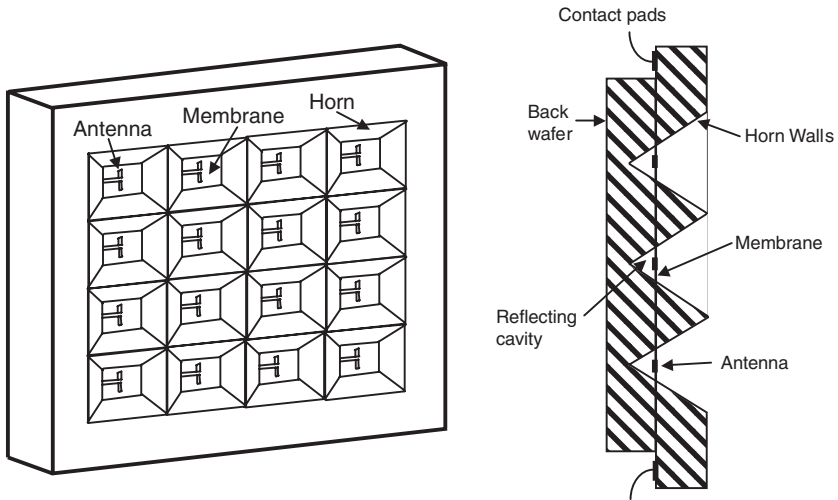


FIGURE 23-12 Two-dimensional array utilizing membranes and air cavities to reduce losses (after G. Rebeiz et al¹³ © IEEE 1990)

The horns were fabricated by anisotropic etching of the silicon in an ethylenediamine-pyrocatechol solution. The etchant used forms pyramidal holes bounded by $\langle 111 \rangle$ crystal planes in $\langle 100 \rangle$ silicon. The flare angle of the horn was fixed at 70.6 degrees by the orientation of the crystal planes. This angle was larger than desired to accommodate the employed fabrication method. It is possible that the horn angle could be better controlled if reactive ion etching or ion-beam milling were used. In this case, the membrane was produced by depositing a silicon oxynitride layer on the front wafer using chemical vapor deposition. The silicon was then etched away to leave the free-standing membrane. It is noted that for the membrane to be flat and rigid, it must be in tension.

An additional potential challenge associated with on-chip antennas is the limitation in dielectric properties and thickness. Silicon substrates have a relative permittivity of about 11.7 but the resistivity of the substrate can vary. Of course, substrates are limited to those available for already established semiconductor processes, and wafer thicknesses may also be limited to stay with standard processing. Typical high-volume silicon substrates may have resistivity on the order of $10 \Omega\text{-cm}$, though higher resistivities of $1 \text{ k } \Omega\text{-cm}$ or more are available but are more expensive and may not be available in large diameters. Of course, substrate thickness and the dielectric constant can effect antenna performance beyond dielectric and ohmic losses. For example, if the substrate thickness and dielectric constant are not chosen properly, surface waves and other substrate modes can significantly effect efficiency and bandwidth¹⁵ (see Figure 23-13). Conditions that generate surface wave and substrate modes can significantly affect performance so they should be avoided.

Antennas implemented at high frequencies tend to be relatively simple geometries and try to maximize metallization to minimize ohmic losses. Figure 23-14 is an example of an edge-fed microstrip antenna that can be implemented in a very small space by trading bandwidth.

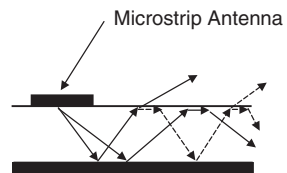


FIGURE 23-13 Formation of surface wave and other substrate modes can effect performance of on-chip antennas (after N. Alexopoulos et al¹⁵ © IEEE 1983)

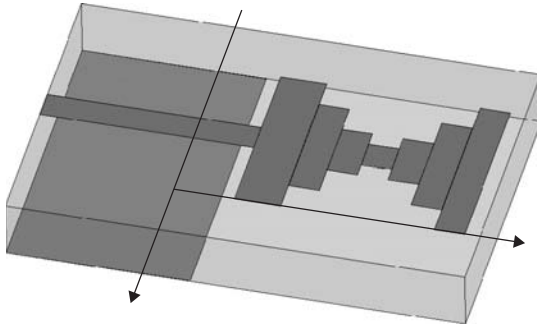


FIGURE 23-14 Example layout of an edge-fed element on silicon for compact on-chip antennas. When minimized in size, the antenna element measures only $360\ \mu\text{m} \times 135\ \mu\text{m}$ delivering 3.2-dBi gain and a 3-dB bandwidth of 17%.

This microstrip antenna can be optimized to occupy only $360\ \mu\text{m} \times 135\ \mu\text{m}$ and deliver a gain of 3.2 dBi with a corresponding 3-dB bandwidth of 17%. Compact antenna elements such as this can be used to realize higher gain by implementing an array of on-chip elements.

Another antenna element that shows promise is illustrated in Figure 23-15. This element is formed from a half circle and incorporates a tuning slot that allows this element to be tuned so it can be impedance-matched more easily. This is another example of a broadband element that can be implemented in a relatively small space in trade for bandwidth. Though this element is larger than the previous example, it can be designed to achieve a 3-dB 17% bandwidth gain of 3.8 dBi while occupying only $480\ \mu\text{m} \times 240\ \mu\text{m}$. Each of these approaches can be implemented with much greater bandwidth if larger physical sizes are acceptable; each of these example microstrip antenna elements offers a wide range of usefulness and can be tailored for a particular application where size, bandwidth, or gain may be the most important design parameter.

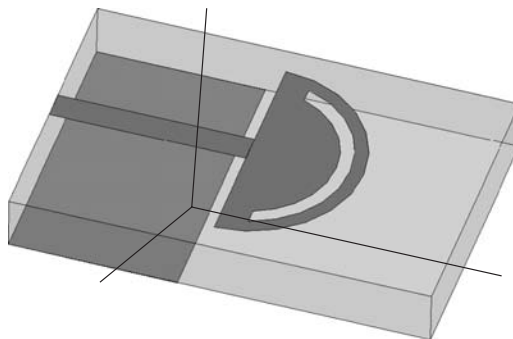


FIGURE 23-15 Example of a millimeter wave antenna comprised of a half-circle element with a tuning slot. When the size is minimized, the element measures $480\ \mu\text{m} \times 240\ \mu\text{m}$ delivering a gain of 3.8 dBi with a 3-dB bandwidth of 17%.

22.3 SUBMILLIMETER-WAVE AND TERAHERTZ ANTENNAS

Approaches used at submillimeter-wave and THz frequencies are very similar to those already discussed. Fundamentally, the fabrication challenges and losses are the same but more of an issue at THz frequencies. At submillimeter wavelengths, fabrication techniques are limited to micromachining and semiconductor lithography or equivalent due to their smaller dimensions. Interconnects between the antenna and other components such as detectors (see Chapter 32) are also more challenging, leading to higher levels of integration to achieve acceptable performance.

An example of a log-periodic antenna fabricated on a thin membrane with an integrated detector operating up to 700 GHz¹⁶ is shown in Figure 23-16. As shown, membranes and air cavities along with an integrated antenna detector were employed to achieve performance. Antennas operating at frequencies greater than 1 THz have also been demonstrated. As an example, a log-periodic antenna element with a lens coupled to a hot-electron bolometer was fabricated and shown to operate from 1–6 THz¹⁷ (see Figure 23-17). Again, lithography requirements are not achievable with standard printed wiring-board techniques, and integration with other components is required to minimize interconnect losses.

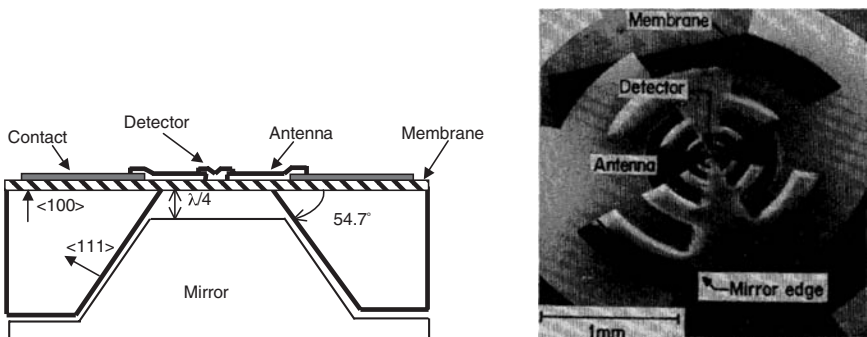


FIGURE 23-16 Example antenna on a thin membrane integrated with a detector operating up to 700 GHz (G. Rebeiz et al¹⁶ © IEEE 1987). As the operational frequency increases, higher levels of integration are required to avoid degradation due to interconnects.

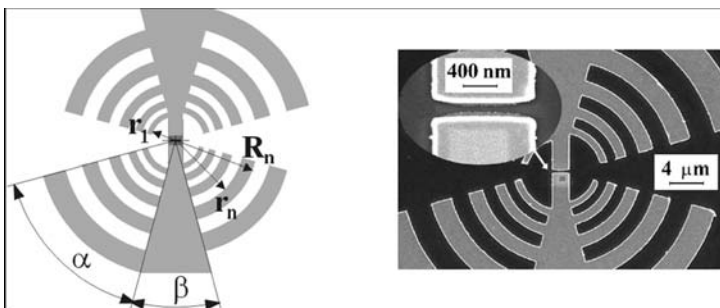


FIGURE 23-17 Log-periodic antenna element with a lens coupled to a hot-electron bolometer for operation at 1–6 THz (after A. Semenov et al¹⁷ © IEEE 2007)

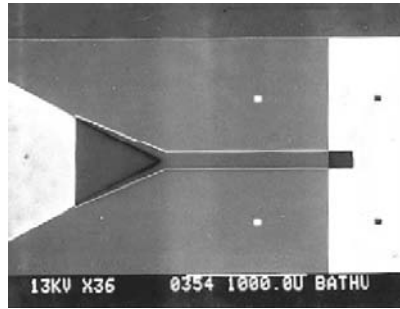


FIGURE 23-18 Micromachined waveguide antenna for 1.6-THz operation (after J. Bowen et al^{18,19} © IEEE 2006). A gain of about 13 dBi was achieved.

Micromachining has also been used to successfully fabricate THz antennas. A 1.6-THz waveguide antenna fabricated using micromachining^{18,19} is shown in Figure 23-18. Though there is some room for improving the performance of this antenna, micromachining techniques show promise for THz antennas. A gain of about 13 dBi was achieved in this prototype. As in other cases, volume fabrication has not been demonstrated, but it does show promise for specialized applications within the THz frequency range. An alternative to micromachining is standard semiconductor lithography techniques to create repeatable feature sizes well into the submicron range.

In addition to the above, other fabrication approaches have also been explored for THz antennas. One such approach is the photoconductive antenna,²⁰ also known as a photoconductive switch and shown in Figure 23-19. In this case, electrodes are formed on a semiconducting photoconductive film. The film is (in most cases) a III-V semiconductor such as GaAs. The photoconductive antenna can be considered as a dipole of length L with its resonance shifted due to the refractive index of the semiconductor (GaAs at THz frequencies has a refractive index of $n = 3.4$). For operation, a voltage is established across the electrical contacts and the excited carriers are accelerated by the electric field during the optical excitation pulse. This results in a short broadband electromagnetic pulse in the THz region. When used as a receiver, during the optical pulse the excited carriers are accelerated by the electric field component of the THz pulse, which creates a measurable current. The receiver is, as usual, connected to a current amplifier across the terminal electrical contacts.

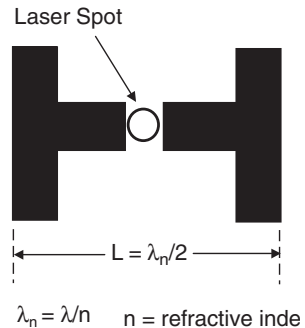


FIGURE 13-19 Photoconductive antenna to generate and transmit or receive THz signals (L is antenna length)

REFERENCES

1. R. Emrick, S. Franson, J. Holmes, B. Bosco, and S. Rockwell, "Technology for Emerging Commercial Applications at Millimeter-Wave Frequencies," *IEEE/ACES International Conference on Wireless Communications and Applied Computational Electromagnetics* (April 2005): 425–429.
2. C. Doan, S. E. Mami, D. Sobel, A. Niknejad, and R. Brodersen, "60 GHz CMOS radio for Gb/s Wireless LAN," *IEEE Radio Frequency Integrated Circuits Conference* (2004): 225–228.

3. R. M. Emrick and J. L. Volakis, "Antenna Requirements Short Range High Speed Wireless Systems Operating at Millimeter-wave Frequencies," *2006 IEEE IMS International Microwave Symposium* (June 2006): 974–977.
4. L. Yujiri, M. Shoucri, and P. Moffa, "Passive Millimeter-Wave Imaging," *IEEE Microwave Magazine*, vol. 4, issue 3 (September 2003): 39–50.
5. P. Siegel, "THz Applications for Outer and Inner Space," *17th International Zurich Symposium on Electromagnetic Compatibility* (February 2006): 1–4.
6. T. Muller, W. Grabherr, and B. Abelsech, "Surface-mountable Metalized Plastic Waveguide Filter Suitable for High Volume Production," *European Microwave Conference*, vol. 3 (October 2003): 1255–1258.
7. J. Hirokawa and M. Ando, "40 GHz Parallel Plate Slot Array Fed by Single-layer Waveguide Consisting of Poast in a Dielectric Substrate," *IEEE Antennas and Propagation Society International Symposium*, 1998.
8. D. Pozar, "Aperture Coupled Waveguide Feeds for Microstrip Antennas and Microstrip Couplers," *IEEE Antennas and Propagation Society International Symposium*, vol. 1 (July 1996): 700–703.
9. K. Maruhashi, M. Ito, L. Desclos, K. Ikuina, N. Senba, N. Takahashi, and K. Ohata, "Low-Cost 60 GHz-Band Antenna-Integrated Transmitter/Receiver Modules Utilizing Multi-Layer Low Temperature Co-Fired Ceramic Technology," *IEEE International Solid-State Circuits Conference* (2000): 324–325.
10. T. Seki, N. Honma, K. Nishikawa, and K. Tsunekawa, "A 60-GHz Multilayer Parasitic Microstrip Array Antenna on LTCC Substrate for System-on-Package," *IEEE Microwave and Wireless Components Letters*, vol. 15, issue 5 (May 2005): 339–341.
11. R. Bairavasubramanian, D. Thompson, G. DeJean, G. Ponchak, M. Tentzeris, and J. Papapolymerou, "Development of mm-Wave Dual-Frequency Multilayer Antenna Arrays on Liquid Crystal Polymer (LCP) Substrate," *IEEE Antennas and Propagation Society International Symposium*, vol. 1B (July 2005): 393–396.
12. C. Nguyen, L. Katehi, and G. Rebeiz, "Micromachined Devices for Wireless Communications," *Proceedings of the IEEE*, vol. 86, no. 8 (August 1998).
13. G. Rebeiz, D. Kasilingam, Y. Guo, P. Stimson, and D. Rutledge, "Monolithic Millimeter-Wave Two-Dimensional Horn Imaging Arrays," *IEEE Transactions on Antennas and Propagation*, vol. 38, no. 9 (September 1990).
14. G. Rebeiz, "Millimeter-Wave and Terahertz Integrated Circuit Antennas," *Proceedings of the IEEE*, vol. 80, no. 11 (November 1992).
15. N. Alexopoulos, P. Katehi, and D. Rutledge, "Substrate Optimization for Integrated Circuit Antenna," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-31, no. 7 (July 1983).
16. G. Rebeiz, W. Regehr, D. Rutledge, R. Savage, and N. Luhmann, "Submillimeter-wave Antennas on Thin Membranes," *IEEE Antennas and Propagation Society International Symposium*, vol. 25, (June 1987): 1194–1197.
17. A. Semenov, H. Richter, H. Hubers, B. Gunther, A. Smirnov, K. Ll'in, M. Siegel, and P. Karamarkovic, "Terahertz Performance of Integrated Lens Antennas with a Hot-Electron Bolometer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 2 (February 2007).
18. J. Bowen, S. Hadjiloucas, B. Towlson, L. Karatzas, S. Wootton, N. Cronin, S. Davies, C. McIntosh, J. Chamberlain, R. Mmiles, and R. Pollard, "Micromachined Waveguide Antennas for 1.6 THz," *IEEE Electronics Letters*, vol. 42, no. 15 (July 2006).
19. J. Bowen, S. Hadjiloucas, B. Towlson, L. Karatzas, S. Wootton, N. Cronin, S. Davies, C. McIntosh, J. Chamberlain, R. Mmiles, and R. Pollard, "Micromachined Waveguide Antennas for 1.6 THz," *IEEE Electronic Letters*, vol. 42, no. 15 (July 2006).
20. H. Murakami, S. Ono, A. Quema, G. Diwa, E. Estacio, N. Sarukura, R. Naqasaka, Y. Ichikawa, E. Ohshima, H. Oqino, A. Yoshikawa, and T. Fukuda, "Zinc Oxide Single Crystal as Substrate for Photoconductive Antenna Device Generating Radiation in the Terahertz Frequency Region," *Infrared and Millimeter Waves and 13th International Conference on Terahertz Electronics*, vol. 2 (September 2005): 547–548.

Chapter 24

Ultra Wideband
Arrays

J. J. Lee

Raytheon Space and Airborne Systems

CONTENTS

24.1 INTRODUCTION.....	24-2
24.2 WIDEBAND ARRAY CONCEPTS	24-2
24.3 OTHER CONNECTED-ARRAY DESIGNS	24-16
24.4 WIDEBAND GROUND PLANE.....	24-18
24.5 PROTOTYPE WIDEBAND ARRAYS	24-20

24.1 INTRODUCTION

There are many reference books on antenna theory and design,¹⁻¹⁷ but very few are devoted to wideband phased arrays. A complete and comprehensive treatment of all aspects of phased arrays is given by Hansen in *Phased Array Antennas*¹⁷ and in his previous works. The array antenna field is considered to be mature and well established; however, recent years have seen a rapid growth in the development of ultra-wideband arrays, primarily driven by the requirements of modern radar and Electronic Warfare (EW) systems that require very wide bandwidth from UHF to high X-band. Also, there is an increasing need to consolidate and reduce the number of antennas on an airborne and shipboard platform due to limited space in accommodating an ever proliferating number of sensors. Furthermore, with the advent of channelized digital receivers backed by a massive amount of computing power, adaptive digital beamformers with unique processing features have become a reality and hence have received wide attention in the defense industry. To achieve these goals, an ultra-wideband aperture capable of supporting more than 10:1 band is desired. A great deal of progress has been made, so this chapter reviews the design concepts of wideband arrays and updates the progress reported by various groups.

Section 24.2 reviews the definition of bandwidth, principles of wideband arrays based on the continuous current sheet model by Wheeler,¹⁸ and the rigorous formulation recently developed for practical array design. Section 24.3 discusses recent progress of wideband arrays using “connected” elements. Section 24.4 examines the ground plane effects, and Section 24.5 presents some prototype examples of wideband arrays, including single-pol (polarization) and dual-pol apertures such as long slot and low profile egg-crate designs.

24.2 WIDEBAND ARRAY CONCEPTS

Definition of Wideband

To begin, we should define a more useful term than conventional “bandwidth” for wideband applications. For all practical purposes, the bandwidth of an array antenna is defined here by the frequency range over which the input VSWR is less than 2:1. Typical (fractional) bandwidth of many existing radar systems is on the order of 10 to 20 percent, with the bandwidth defined as

$$\text{Bandwidth (\%)} = 200 \frac{f_2 - f_1}{f_2 + f_1} \quad (24-1)$$

where f_1 is the low end and f_2 the high end of the band. For wideband systems, it is common to specify the ratio of the high end and low end frequency with a factor N , where $N = f_2 / f_1$. For instance, a UHF radar antenna designed for foliage penetration (FOPEN) to detect targets concealed under forests may require a bandwidth ranging from 150 to 600 MHz, which is 4:1 band. On the other hand, a typical high-band EW array may require 3:1 band from 6 to 18 GHz. In such cases, it is more convenient to use the N :1 ratio in the discussion of system requirements and performance. The conversion between the fractional bandwidth and the N factor is given by

$$\text{Bandwidth (\%)} = 200 \frac{N - 1}{N + 1} \quad (24-2)$$

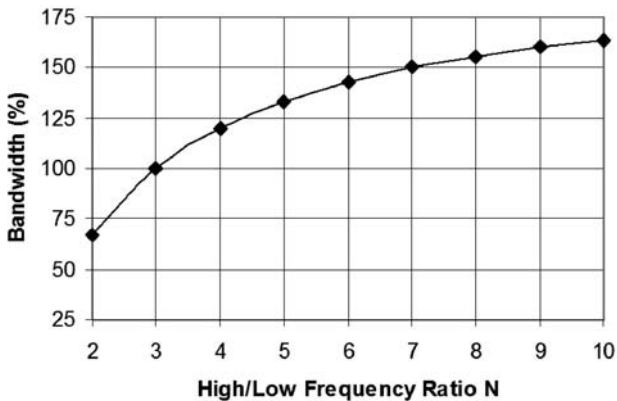


FIGURE 24-1 Fractional bandwidth vs. frequency ratio factor N :1

The percent bandwidth for some typical N factor is shown in Figure 24-1. The fractional bandwidth is meaningful for narrowband systems but not so useful for the wideband cases. For instance, a very challenging 10:1 band translates into a bandwidth of 164 percent, while an extremely wideband system ($>100:1$) has 200 percent fractional bandwidth as a limit. Hence, the N factor is more suitable for describing a wideband system. A large N :1 ratio usually signifies great difficulty in providing the wide bandwidth for a mission, because the array antenna must be able to support a system where the wavelength varies by a factor of N . It also implies that the array lattice is tight, and the element spacing is limited to about 0.6 wavelength at the high end of the band for grating-lobe-free operation over a wide scan ($\pm 60^\circ$) volume.

We may ask, “how wideband is wideband?” The answer lies in the system requirements and the center frequency of the operation. Depending on applications, 20 percent bandwidth at X-band (10 GHz) is 2 GHz, which may be sufficient to provide adequate range resolution for detection, imagery, and mapping. For more advanced systems, a 2:1 band from 6 to 12 GHz may be more desirable. For direction finding, surveillance, and certain EW applications, a 3:1 band may be required. For low frequency such as UHF, a mere 500-MHz bandwidth centered at 350 MHz implies that the array must operate over a 6:1 band from 100 to 600 MHz. Thus it is harder to get wide instantaneous bandwidth at low frequency. A bowtie element at 400 MHz is on the order of 30 cm, while at 100 MHz it increases to 120 cm. This natural constraint in wavelength is a critical issue to be considered in the design of a wideband array.

While an ultra-wideband aperture offers many desirable features, it is not without penalty and drawbacks. The price to pay is in higher component costs for the RF front-end and in complexity in the dense packaging of the feed networks. The lattice designed for the high end of the band characterized by small element spacing is wasteful for the low end. To support wide band, the feed components including transitions and baluns require extra space and multiple layers of circuit board to support the bandwidth. The extra RF loss associated with these front-end components must be traded off with the benefits gained from the extended bandwidth. Sometimes the bottleneck of a wideband array is not the radiating aperture, but the T/R modules and the associated electronics that must be packed to fit in a tight lattice and to handle the intense heat dissipation in an array environment.

It should be emphasized that a phased array is a periodic structure characterized by certain boundary conditions and eigen (Floquet) modes. A wideband radiating element

does not guarantee that the array will be broadbanded, but a high- Q narrowband (resonant) element will definitely make it difficult for the array to operate over a wide band. Aperture matching is a complicated subject that is configuration and radiator-geometry dependent. For instance, Amitay et al¹⁴ devoted their whole book just to analyzing the waveguide array using eigen mode expansion and matching techniques. However, by and large if the radiating element and the lattice structure do not excite a large number of reactive modes, it is more likely to match the aperture impedance and transmit the power efficiently. In other words, the more an aperture field excited over the array resembles a plane wave, the easier it is to radiate its power into the free space over a wide band.

Current Sheet Model

The basic concept of aperture matching can be dated back to 1948 when Wheeler¹⁸ first discussed the radiation resistance of an infinite array. Later he described the concept of a continuous current sheet model associated with a hypothetical array radiating a TEM wave into the free space.¹⁹ For broadside radiation of any frequency, the radiation resistance of an idealized current sheet backed by a ground plane offset by $\frac{1}{4}\lambda$, is 120π (377) Ω . Conversely, a resistive plane with a surface impedance equal to the characteristic impedance of the plane wave (377 Ω), backed by a ground plane separated by $\frac{1}{4}\lambda$, will absorb an incident plane wave without reflection. The impedance is reduced to 60π (188) Ω if the current sheet is radiating into both directions without the ground plane. For E -plane scan, with an oblique incident angle θ from broadside, the boundary resistance decreases, proportional to $\cos \theta$, due to the cosine projection of the E -field on the aperture (reduction in voltage). For H -plane scan, the boundary resistance increases and varies as $1/\cos \theta$, due to the cosine projection of the H -field on the aperture (reduction in current density). In reality, such an infinite and uniform current sheet does not exist. A more practical case close to the ideal situation is an infinite array of closely spaced conducting strips carrying continuous current over the aperture, as shown in Figure 24-2. An approximate way to produce such a current distribution is to drive each long wire or strip with multiple feed points, effectively exciting an infinite array of connected short dipoles, as depicted in Figure 24-3. The element spacing in both directions must be small, on the order of 0.6λ at the high end of the band, so that it keeps the grating lobes out of the visible space over the desired frequency range and scan volume. The task now is to determine the input impedance of such an array, and to provide a matching network with baluns and transformers to efficiently radiate the power into the free space with minimum reflection over a wide band.

The subject of conducting wires and linear connected arrays has been studied by Hansen,²⁰⁻²⁴ who used the Galerkin moment method to analyze the current distribution over a finite array of such elements. He reported that for arrays of halfwave dipoles the current exhibits strong resonances along the array, with a period of one wavelength. This is apparently due to the edge currents present in finite dipole arrays not being attenuated in connected arrays. For dipole lengths much shorter than halfwave, the current is roughly constant over the entire array. This results in an approximation to the Wheeler's current sheet model, and input impedance ranges from 200 to 400 Ω . Overall, a connected dipole array provides better wideband gain than a corresponding unconnected dipole array.

Characteristics of Wideband Array

Significant insights into the linear connected dipole array can be found in the works by Neto and Lee,²⁵⁻²⁷ who studied the characteristics of an infinite 2D long slot array shown

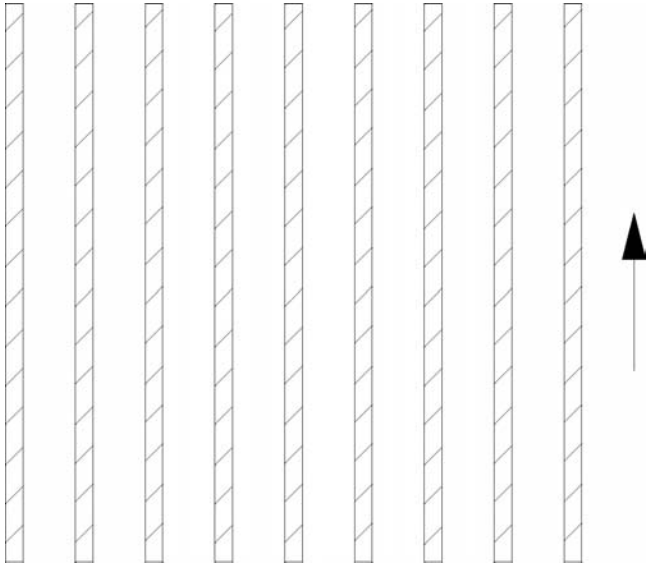


FIGURE 24-2 An array of conducting strips carrying constant current resembles Wheeler's current sheet model of a wideband array.

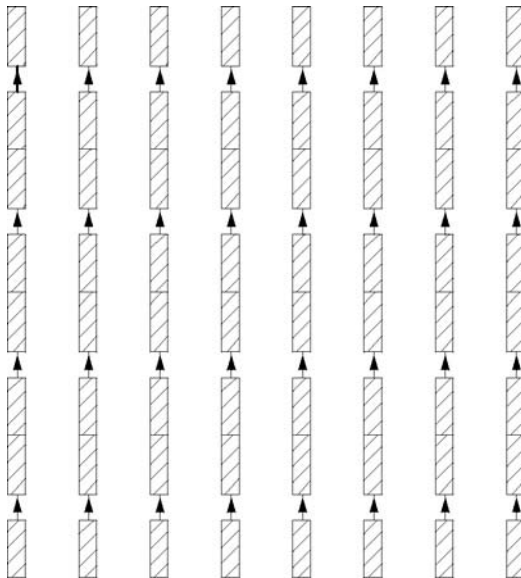


FIGURE 24-3 An array of long conducting strips with multiple feed points can produce a quasi-uniform current sheet. Effectively, it is an array of connected short dipoles excited by multiple electric current delta sources.

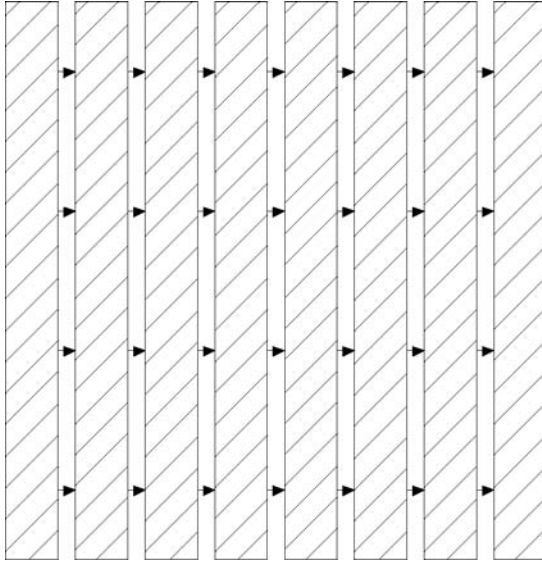


FIGURE 24-4 An array of long slots with multiple feed points is a conjugate configuration of the connected dipole array.

in Figure 24-4. Each long slot is excited by multiple delta current sources, equally spaced at about 0.6λ with respect to the high end of the band to avoid grating lobes. Since the long slot array is a conjugate (complementary) version of the connected dipole array, it is an array of connected magnetic dipoles, as shown in Figure 24-5. The duality of these two arrays becomes evident when Figure 24-3 is compared with Figure 24-5. Babinet's principle^{28,29} states that the characteristics of these two arrays are identical except that the E - and H -field are interchanged in the radiation. Neto's rigorous analysis of the long slot array provides a solid background to understanding why a phased array with features resembling Wheeler's current sheet model is intrinsically wideband. The treatment is general and applicable to a broad range of related designs including connected dipole arrays. The theory provides physical insights into the characteristics of a wideband array with connected elements. Thus it is beneficial to review the analytic formulation and to identify the key steps in the derivation for better understanding of the subject. This summary is provided for handy reference.

Figure 24-6 shows the geometry of an infinite long slot array in free space without any back (ground) plane. The x -oriented slots are etched out in a conducting plane between two homogeneous dielectric half-spaces of the same permittivity ϵ . The slots are centered at $y_n = n_y d_y$, and the cross section (gap) width of each slot, w_s , is uniform in x and is small in terms of a wavelength at the maximum frequency of interest. Each long slot is excited by an array of y -oriented electric delta current sources (generators) of the same length w_s placed at periodic locations. Each delta source is assumed to be thin, but its actual width, t , in x -direction will be considered later. By invoking the equivalence principle, the slot region can be covered by a metallic surface infinitely thin and perfectly conducting, with an unknown magnetic current distribution $\mathbf{m}_{n_y}(x', y') (= -n \times E / \mu_o)$ including the delta sources just above and below the slot region. These magnetic currents have equal

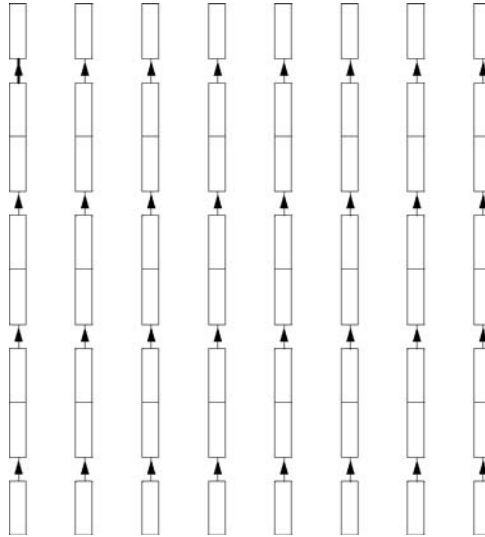


FIGURE 24-5 A long slot array is effectively an array of connected magnetic dipoles excited by multiple magnetic delta current sources.

amplitudes but opposite signs to ensure continuity of the tangential magnetic field through the screen occupied by the slots. Due to the small gap (w_s) of the slots, the magnetic currents over the slots can be assumed to flow in x direction only (first order of approximation). To derive the unknown magnetic current distribution, an integral equation is introduced to enforce the continuity of the total tangential magnetic field above and below the long slot. The total tangential magnetic field H_t on each side of the slot is the sum of the impressed

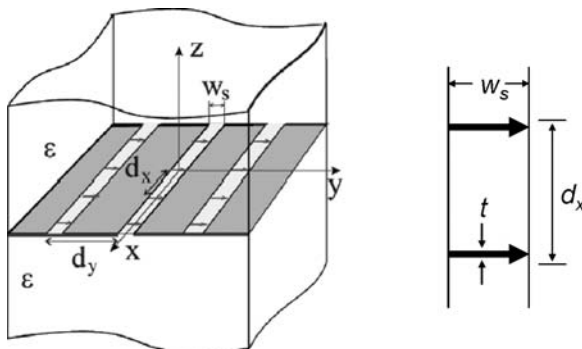


FIGURE 24-6 Geometry of an infinite 2D array of long slots excited by multiple feeds spaced at a Nyquist interval, half wavelength at high end of the frequency band of interest

(forced) field H_f and the scattered field H_s radiated by the magnetic current induced on the slots. This leads to

$$H_s^1 - H_s^2 = H_f^2 - H_f^1 \quad (24-3)$$

where 1 denotes region 1 below the conducting screen (slots) and 2 denotes region 2 above the screen. The right-hand side of Eq. 24-3 represents the discontinuity in the tangential magnetic field produced by the impressed current sources, and it is equal to an infinite series of the delta function by Ampere's law. On the left-hand side H_s^1 and H_s^2 are equal but have opposite signs.

The boundary condition yields an integral equation given by

$$\iint_S g(x - x', n_{y0} d_y - y') m(x', y') dx' dy' = \sum_{n_x=-\infty}^{\infty} f_{n_x}(x) \quad (24-4)$$

with

$$f_{n_x}(x) = I_e \delta(x - n_x d_x) \cdot e^{-jk_{x0} n_x d_x} e^{-jk_{y0} n_{y0} d_y} \quad (24-5)$$

$$m(x', y') = \sum_{n_y=-\infty}^{\infty} m_{n_y}(x', y')$$

where S represents the ensemble of the slots, k_{x0} , k_{y0} are the phase shifts of the impressed excitation, I_e along x and y , respectively,

$$k_{x0} = k_0 \sin \theta \cos \phi$$

$$k_{y0} = k_0 \sin \theta \sin \phi$$

and $g(x, y)$ is the Green's function of the magnetic field due to a delta magnetic current element. In Eq. 24-5 $m_{n_y}(x', y')$ represents the unknown magnetic current distribution along the n_y -th long slot located at $x_n = n_y d_y$. The left-hand side of Eq. 24-4 represents the difference between the magnetic fields at $x(y = 0, z = 0)$ along the zero-th (central) long slot, radiated by the magnetic currents above and below the slots. The procedure to solve the unknown $m_{n_y}(x', y')$ in Eq. 24-4 is similar to the treatment given by Elliott³⁰ where a moment method is applied to find the current distribution on an electric dipole. The long slot problem can be solved by treating the slot element as a magnetic dipole. In the formulation the total magnetic field radiated by all the slots including the delta sources (generators) at a point in space (x, y, z) is expressed as an integral of the Green's function weighted by the unknown magnetic current distribution $m_{n_y}(x', y')$ from all the slots. The boundary condition is then imposed along the zero-th (central) long slot, requiring that the total tangential magnetic field H_x vanish on the perfect magnetic conductor²⁸ occupied by the slot region except at the gap of each delta source in Figure 24-5. This condition leads to Eq. 24-4. Next, the magnetic current distribution $m_{n_y}(x', y')$ is assumed to consist of two separable parts: a longitudinal x -dependence $v(x')$ and (transverse) y -dependence $m_t(x', y')$. It is the longitudinal part of the magnetic current distribution $v(x')$ that we want to determine, so that the voltage across the slot at x can be computed from the E -field by the relation of

$$m_{n_y}(x', y') = v(x') m_t = -n \times E$$

Then the integral equation is transformed into the spectral domain k_x, k_y on both sides to isolate and find the Fourier components of the magnetic current distribution. The spectral component of the free space Green's function for the magnetic field radiated by a delta magnetic current element (in the x direction) is derived separately, given by

$$G_{xx}^{fs} = -\frac{1}{k_0 \zeta_0} \frac{1}{2} \frac{k_0^2 - k_x^2}{\sqrt{k_0^2 - k_x^2 - k_y^2}} \quad (24-6)$$

This expression is identical to the Green's function for the electric field radiated by a delta electric current element (in the x direction) derived by Lee,³¹ except the ground plane is removed, and $\sin(k_z z')$ is replaced by the decaying exponential term with z' set to zero. In the inverse transform to find $v(x')$, the Poisson sum formula is used in the integration of k_x, k_y to obtain a double series contributed by the discrete Floquet modes in both x and y directions. This is the nature of the periodic structure. These modes represent the spectral lines extracted from the continuum (k_x, k_y) spectrum associated with an isolated single slot element (magnetic dipole). The result for the voltage along the central long slot at $y = 0$ is given by

$$v_0(x) = \frac{d_y}{d_x} \frac{k_0 \zeta_0}{2} \sum_{m_x=-\infty}^{\infty} \frac{I_e e^{-jk_{xm}x}}{(k_0^2 - k_{xm}^2) \sum_{m_y=-\infty}^{\infty} \frac{J_0\left(k_{ym} \frac{w_s}{2}\right)}{\sqrt{k_0^2 - k_{xm}^2 - k_{ym}^2}}} \quad (24-7)$$

where k_0 and ζ_0 are the free-space wave number and characteristic impedance, (120π) 377 Ω , and

$$k_{xm} = k_{x0} - 2\pi m_x / d_x$$

$$k_{ym} = k_{y0} - 2\pi m_y / d_y$$

are the wave numbers of the Floquet modes in the x and y direction. The Bessel function $J_0(k_{ym} w_s / 2)$ results from the integration of the magnetic current across the slot in the y direction, a factor associated with the effects of the slot width. From the voltage expression (see Eq. 24-7), the active impedance of the central slot element at $x = 0$ with normalized input current $I_e = 1$ is

$$z_{in} = \frac{d_y}{d_x} \frac{k_0 \zeta_0}{2} \sum_{m_x=-\infty}^{\infty} \frac{\text{sinc}^2\left(\frac{k_{xm} t}{2}\right)}{(k_0^2 - k_{xm}^2) \sum_{m_y=-\infty}^{\infty} \frac{J_0\left(k_{ym} \frac{w_s}{2}\right)}{\sqrt{k_0^2 - k_{xm}^2 - k_{ym}^2}}} \quad (24-8)$$

where the sinc function results from the integration of the constant magnetic field across a finite gap (t) at the source (see Figure 24-5), accounting for the effects of the feed probe's finite width (see Figure 24-6). The factor $J_0(k_{ym} \frac{w_s}{2})$ stems from the edge condition of the current source across the slot.

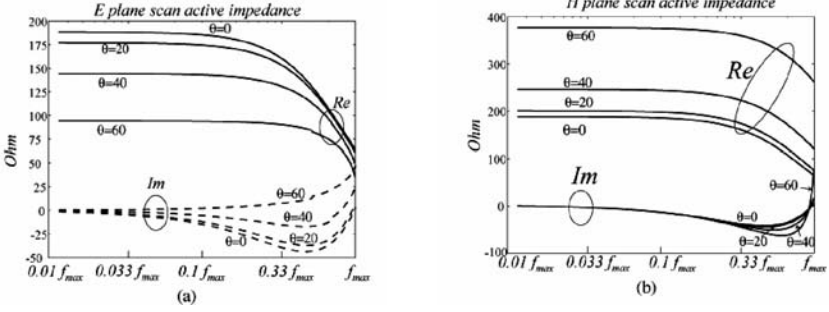


FIGURE 24-7 Input impedance of an infinite long slot array as a function of frequency and scan angle in (a) E -plane and (b) H -plane for $d_x = d_y = 0.5\lambda_{\min}$, $\lambda_{\min} = c/f_{\max}$, $w_s = 0.1\lambda_{\min}$, $t = 0.1\lambda_{\min}$

Now the characteristics of a wideband array with connected elements, be they electric or magnetic dipoles, can be discussed in detail. Figure 24-7 displays the input impedance of the array with all the mutual coupling effects taken into account when scanned in the E - and H -planes. One striking feature of such an array is that for frequencies far below the Nyquist limit (when the element spacing dx approaches half wavelength at the high end of the band), the impedance is almost constant, mostly resistive with very little reactance. There is practically no cutoff at the low end of the band for an infinite array without a back plane. The impedance is $(60\pi) 188 \Omega$ at broadside, equal to half the characteristic impedance of the free space, because the array is radiating bi-directionally into two half spaces. The low-frequency characteristics of the array can be obtained by retaining only the lowest-order terms ($m_x = 0$, $m_y = 0$) in Eq. 24-8, leading to a frequency-independent impedance

$$z_{\text{in}} \approx \frac{d_y}{d_x} \frac{\zeta_0}{2} \frac{\cos \theta}{1 - \sin^2 \theta \cos^2 \phi} \quad (24-9)$$

where the Bessel function and the sinc function have been set to unity for small arguments. This first order approximation validates Wheeler's current sheet model, i.e., the scan impedance decreases from 188 to 94 Ω , proportional to $\cos \theta$, in the E -plane scan ($\phi = 90^\circ$), but increases from 188 to 377 Ω proportional to $1/\cos \theta$ in the H -plane scan ($\phi = 0^\circ$). This confirms that an array designed to carry quasi-uniform current distribution over the aperture can operate over an ultra wideband in the sense that the fractional bandwidth $N = f_{\text{high}}/f_{\text{low}}$ approaches infinity for a large array. The scan performance of the long slot array is summarized in Figure 24-8, showing the reflection coefficient of the radiating element as a function of the scan angle θ in E - and H -planes. These curves indicate that the long slot array is well behaved for 2D scan over a wide band and scan volume, as long as the frequency is not too close to the upper limit defined by the Nyquist interval when the element spacing in x and y approaches one half wavelength. Note that Z_{in} is not sensitive to the width of the slot (w_s) or the source probe (t). They do not play a significant role until the frequency approaches the high end of the band. Referring to Figure 24-5 of the equivalent magnetic dipole array, a wider slot width means fatter dipoles with wider bandwidth. Similarly, a wider strip of t (thicker probe at the excitation point) tends to decrease the inductance of the source element. It is the ratio of w_s and t that we need to adjust to cancel the reactance at the feed port.

The imaginary part of the input impedance comes from the square root of Eq. 24-8 associated with the modes of $|m_x|$ and $|m_y| > 0$. These higher order modes are

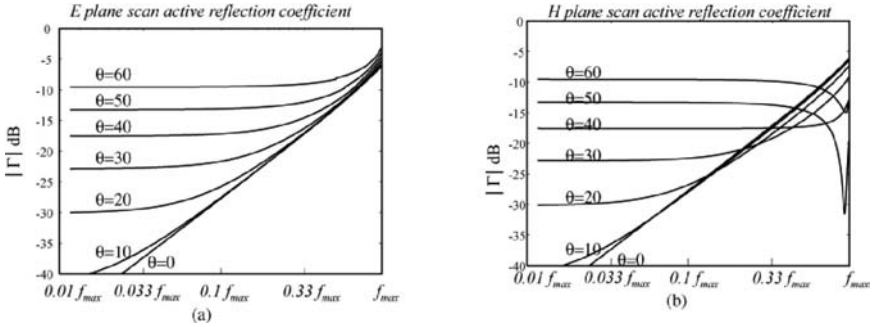


FIGURE 24-8 Reflection coefficient of the slot element normalized to an input feed impedance of 188 Ω for scan in (a) *E*-plane and (b) *H*-plane

associated with the reactive power stored in the local source region. When the slot is widened to increase the bandwidth, the gap source (excitation probe) also needs to be widened to minimize excessive inductance at the input for high frequency end. An example of achieving this is presented in Lee et al,³² where the probe is split into two branches to feed the slot.

Ground Plane Effects

The infinite array discussed so far has no back plane, and it radiates bi-directionally. For all practical purposes, however, there is always a metallic back (ground) plane (see Figure 24-9) to reflect the power for maximum forward gain and to shield the aperture from the electronics mounted on the back plate. This back plane significantly limits the bandwidth of the array. The effects can be analyzed by modifying the Green's function in region 1 between the slot aperture and the back plane. In this case, the Green's function in region 1 is given by

$$G_{xx}^1 = -2j \cot(k_z h) G_{xx}^{fs} \quad (24-10)$$

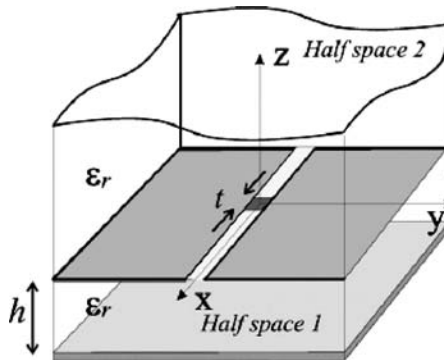


FIGURE 24-9 Unit cell geometry of an infinite 2D slot array backed by a reflecting ground plane

Imposing the boundary condition again, i.e., the total tangential magnetic field contributed from side 1 and side 2 to vanish on the perfect magnetic conductor, we can derive a new formula for the input impedance of the array backed by a ground plane²⁷

$$z_{\text{in}\infty}^{\text{br}} = \frac{k_0 \zeta_0 d_y}{d_x} \sum_{m_x=-\infty}^{\infty} \frac{\text{sinc}^2(k_{xm})}{(k_0^2 - k_{xm}^2)} \times \frac{1}{\sum_{m_y=-\infty}^{\infty} \frac{J_0\left(k_{ym} \frac{w_s}{2}\right)}{k_{ym}} (1 - j \cot(k_{zm} h))} \quad (24-11)$$

As expected, the response is now frequency dependent. When the back plane is about 1/4 wavelength (midband) from the aperture, the back plane presents high impedance to the slots, causing total reflection and maximizing the forward gain. Figure 24-10 shows the complex input impedance of the array, peaking at (120π) 377 Ω around midband with stated array parameters. The analytic results (solid lines) agree well with those obtained by commercial HFSS code (dashed lines). The reflection coefficient plotted in Figure 24-11 shows the bandwidth of the long slot array element connected to a transmission line of 377 Ω characteristic impedance (solid curve). A bandwidth of 47 percent is achieved with 10 dB input match. However, the bandwidth can be improved to 95 percent (3:1 band) by widening the slot width to $0.2 \lambda_{\text{min}}$, changing the back plane distance to $0.4 \lambda_{\text{min}}$, and matching to a 250 Ω input impedance. Arrays with 4:1 bandwidth have been built and tested.^{33,34} The effects of the back plane can be minimized by loading the region in between with resistive or lossy ferrite materials at the expense of efficiency. In fact, an ultrawide bandwidth can be achieved if up to 3 dB loss can be tolerated for some EW applications. In theory, a very compact ultrathin array may be designed to provide greater than 20:1 bandwidth by using next generation meta-ferrite composite materials.³⁵

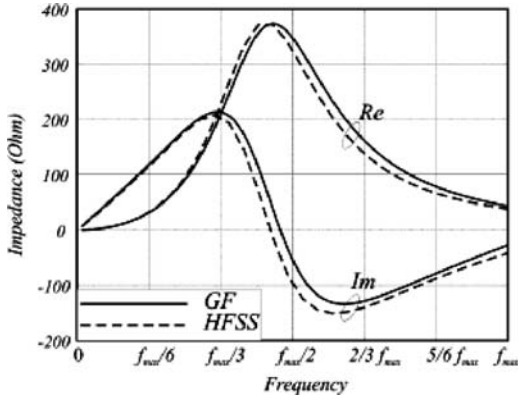


FIGURE 24-10 Input impedance of an infinite long slot array with a back plane. Solid curves are based on Green's function formulation, dotter curves—computed by HFSS unit cell model. ($d_x = d_y = 0.5 \lambda_{\text{min}}$, $\lambda_{\text{min}} = c / f_{\text{max}}$, $w_s = 0.1 \lambda_{\text{min}}$, $t = 0.1 \lambda_{\text{min}}$, $h = 0.25 \lambda_{\text{min}}$.)

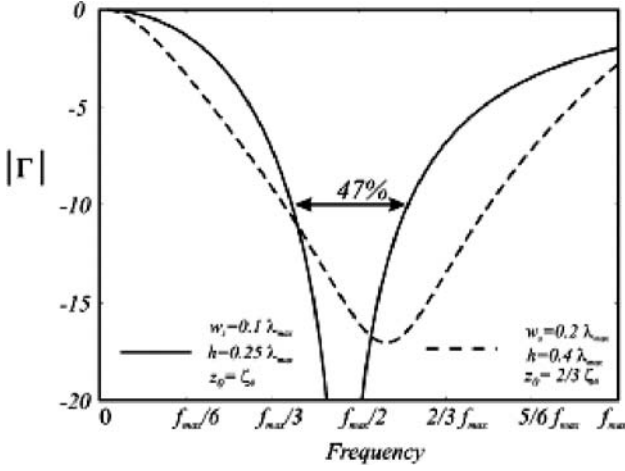


FIGURE 24-11 Reflection coefficient of an infinite long slot array with a back plane (same configuration as Figure 24-9). Bandwidth can be improved to 95 percent (dotted curve) with slightly different array parameters.

Effects of Dielectric Loading

Instead of using a metallic back plane to reflect the power, the array's forward gain can be improved by covering the long slot aperture with a high- κ dielectric radome. Since the dielectric slab is necessarily of finite thickness, reflection will occur at the dielectric-air interface. However, it is well known that a very large bandwidth with negligible reflection can be achieved by using a number of matching layers. For simplicity, we assume that the first dielectric is thick enough to allow the dominant propagating mode to reach the first dielectric interface, and the rest of the matching layers can render the reflection diminishingly small. In this case, as shown in Figure 24-12, the long slot aperture is sandwiched between two half spaces with permittivity of ϵ_{r1} , ϵ_{r2} , respectively. The boundary condition requires that the total tangential magnetic field in the slot, contributed by the magnetic current below and above the slot, vanish. Mathematically, this entails combining the Green's function for region 1 and 2,

$$G(k_x, k_y) = -\frac{1}{k_0 \zeta_0} \left(\frac{k_1^2 - k_x^2}{\sqrt{k_1^2 - k_x^2 - k_y^2}} + \frac{k_2^2 - k_x^2}{\sqrt{k_2^2 - k_x^2 - k_y^2}} \right) \quad (24-12)$$

where $k_1 = k_0 \sqrt{\epsilon_{r1}}$, $k_2 = k_0 \sqrt{\epsilon_{r2}}$. Using this relation, we obtain the input impedance of the array element with dielectric loading given by

$$z_{in}^{2, \text{diel}} = \sum_{m_x=-\infty}^{\infty} \frac{k_0 \zeta_0 d_y \text{sinc}^2(k_{xm} t/2)}{d_x \sum_{m_y=-\infty}^{\infty} \sum_{i=1}^2 \frac{k_i^2 - k_{xm}^2}{\sqrt{k_i^2 - k_{xm}^2 - k_{ym}^2}} J_0\left(k_{ym} \frac{w_x}{2}\right)} \quad (24-13)$$

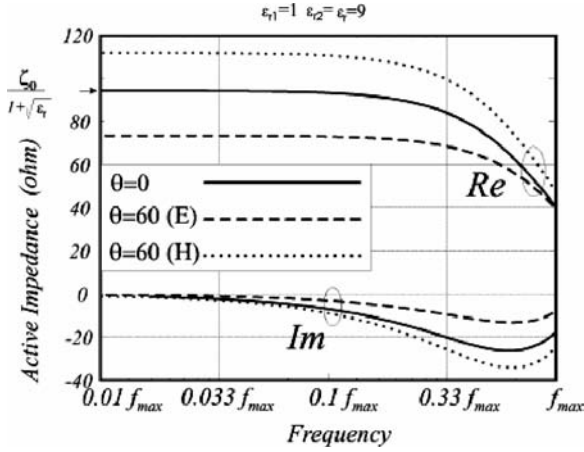


FIGURE 24-12 Active impedance as a function of the frequency for an array of long slots between two dielectrics with $\epsilon_{r1} = 1$, $\epsilon_{r2} = 9$ for broadside radiation, E -plane scan (dashed lines), and H -plane (dotted lines)

As a special case, let region 1 be air and region 2 be loaded with a high- κ material of $\epsilon_r = 9$. Now the impedance of region 2 is reduced by a factor of 3, and hence the front-to-back ratio of the power radiated above and below the aperture is equal to 3:1. Also, the input impedance of the array for broadside radiation is lower, equal to $z_o = 120\pi / (1 + \sqrt{\epsilon_{r2}})$, about 94 Ω for this example, due to the fact that the two regions (like transmission lines) are connected in parallel to the input source. The impedance versus frequency as a function of the scan angle in E - and H -plane is shown in Figure 24-12.

The bandwidth of this radome loaded array will eventually be determined by the number of matching layers employed for a desired front-to-back ratio.

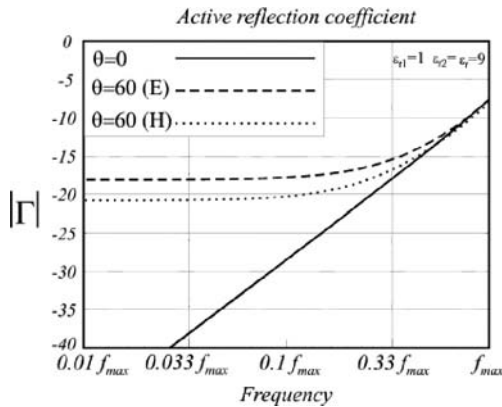


FIGURE 24-13 Reflection coefficients of the array referred to a characteristic input impedance of 94 Ω with $\epsilon_{r1} = 1$, $\epsilon_{r2} = 9$, $z_o = 120\pi / (1 + \sqrt{\epsilon_{r2}})$

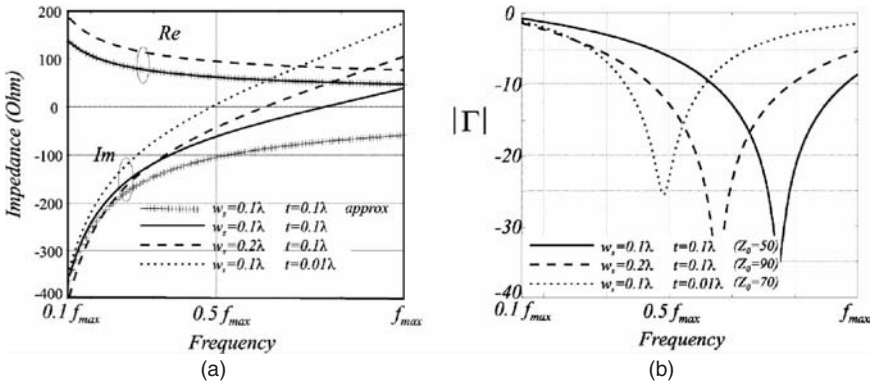


FIGURE 24-14 (a) Input impedance vs. frequency of a single long slot array as a function of slot parameters. Curve with crosses refers to the low-frequency approximation. (b) Reflection coefficient of the array when connected to 50-, 70-, and 90- Ω input line impedance. λ in the figures refers to λ_{min} at maximum frequency.

Finite Array Edge Effects

The wideband array concepts presented so far are based on the theory of an infinite 2D slot array, but all practical designs are finite in dimensions. The truncation effects associated with a finite number of long slots are very similar to those found in standard array antennas.²⁷ The edge effects can be seen by examining a special case where there is only one single long slot in the ground plane. In this case, the edge effects are manifested by a significant amount of reactance in the element's input impedance. The real and imaginary parts of Z_{in} versus frequency are illustrated in Figure 24-14a for various combinations of slot parameters, and the reflection coefficient of the array normalized to various transmission line impedances is shown in Figure 24-14b. Note that the single line source radiates a field that undergoes cylindrical spreading, i.e., the field decays in proportion to $1/\sqrt{\rho}$, with

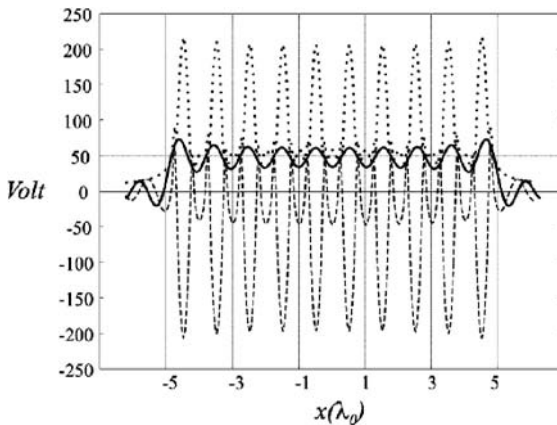


FIGURE 24-15 Voltage distribution on a single infinitely long slot with 20 feed points. The continuous, dashed, and dotted lines represent the real part, the imaginary part, and the amplitude, respectively. $w_s = 0.1$, $t = 0.1\lambda$.

ρ being the distance from the line source. Although it has reasonable bandwidth, its impedance varies significantly at low frequencies. In this limit, with the linear array scanned in the H -plane (along the slot axis), note several features: (a) both real and imaginary parts of the impedance grow as $1/f$, (b) the real part of the impedance is dominated by the slot's width (dipole width), and (c) the impedance is inversely proportional to the cosine of the scan angle. These drawbacks diminish when additional long slots are added in the other direction to form a 2D large array.

On the other hand, the edge effects will show up in the other direction when only a finite number of feed points are in a long slot. As an example, Figure 24-15 shows a case of 20 feed elements in a single slot, where the truncation causes voltage to oscillate around the solution of the infinite array case. This finding is similar to that reported by Hansen²⁴ for a similar finite array configuration.

24.3 OTHER CONNECTED-ARRAY DESIGNS

The idea of connecting the radiating elements in a planar array was apparently first discussed by Baum et al³⁶⁻³⁸ and Inagaki.³⁹ Baum et al studied arrays of square bowtie elements in a complementary configuration. More recently Kesler et al⁴⁰ and Friederich et al⁴¹ developed the “fragmented” element, which comprises a matrix of conducting and dielectric pixels in the unit cell. The fragmented element aperture is a computer-optimized planar system based on the genetic algorithm. Key to the new design was taking advantage of the mutual coupling of the elements. A prototype 33:1 band array (36 cm \times 36 cm) was demonstrated for 0.3–10 GHz where the aperture includes three layers of metal foil fabricated in computer-designed patterns using printed circuit board technology. Layer 1 is the driven layer of radiating elements connected to the input feed ports, while layers 2 and 3 are parasitic layers with dielectric substrates in between for matching to the free space. In addition to the patterned radiating layers, the antenna is backed by six layers of resistive loading material to eliminate reflections from the back plane. The penalty of the lossy backside is 2–3-dB gain drop over the band.

Another connected-element design was presented by Munk.⁴² As shown in Figure 24-16, it is basically an array of short dipoles connected by capacitors. Since the short dipole has very low inductance, the idea is to load down each short dipole with sufficient capacitance to make it resonate and to use the capacitance to tune out the reactance of the ground plane. The resonance and impedance matching concept is a departure from the wideband continuous current sheet model described by Wheeler,¹⁹ and hence cannot be analyzed by the formulation covered earlier in the “Current Sheet Model” subsection. At very low frequency the capacitor-coupled array may not perform satisfactorily, because the reactance of the capacitor becomes large, thus isolating the dipoles from behaving as connected elements to support continuous currents.

Lastly, another kind of array can be considered as one form of the connected element design. It makes use of the well-known Vivaldi element, which has been extensively studied.⁴³⁻⁴⁸ Figure 24-17 shows a generic Vivaldi design. The quasi-end-fire nonresonant radiator is also referred to as a flared notch or tapered slot element. It is typically fed by a “hockey-stick” balun. The feed point flares from a 50- Ω gap to a large opening, matching to free space's wave impedance of 377 Ω in a square lattice. The tapered slot is essentially a transformer that transfers the power to the tip of the elements where the E -field merges into a plane wave. The element is normally fabricated by cutting a notch in a metal plate and backed by a quarter-wave cavity behind the feed point to improve its forward gain. In most cases, the tapered slot element operates in a region where the element is greater than a halfwave long at the low end of the band, and more than two

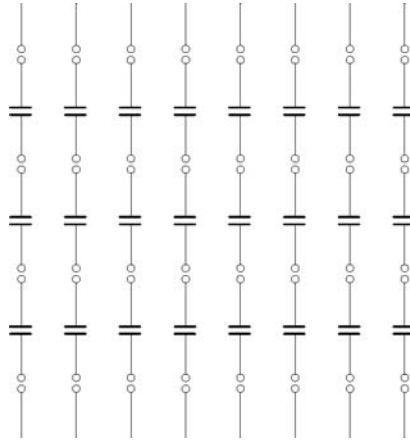


FIGURE 24-16 A linear array of short dipoles connected by capacitors, which are inserted to resonate the inductive dipole elements and tune out the ground plane effects

wavelengths at the high end. These elements may or may not touch each other, depending on the element's length, thickness, and local feed geometry of a specific design. Each design must be modeled by custom software such as HFSS, FEKO, or CST to optimize its overall performance. A long tapered slot radiator is not sensitive to the ground plane effect, and it normally supports 4:1 band but can be designed to cover up to 10:1 bandwidth with some gain ripples.

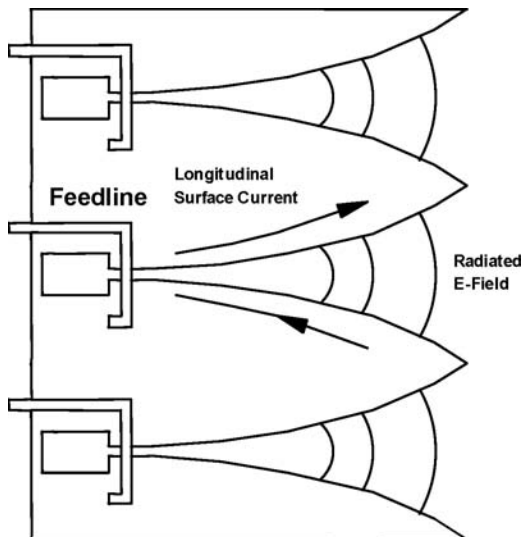


FIGURE 24-17 Tapered slot Vivaldi element typically is $2\text{--}3\lambda$ long at the high end of the band. It supports more than 3:1 band when properly designed.

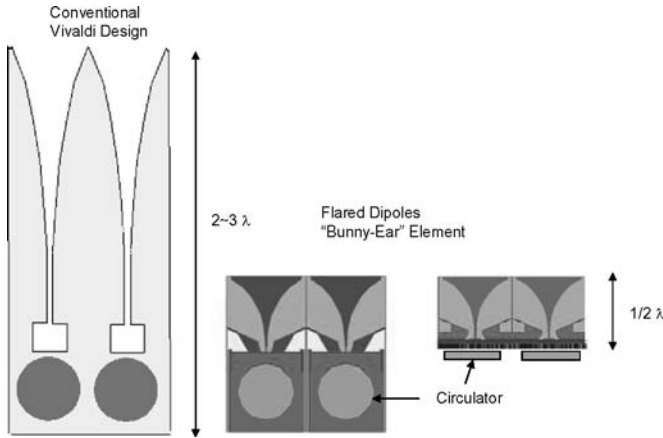


FIGURE 24-18 Low profile flared dipole (bunny-ear) element is typically is $1/2\lambda$ long at the high end of the band. It produces lower cross-pol in the diagonal plane.

Although the tapered slot has been widely used for various applications, it has some drawbacks. First, the longitudinal currents flowing on the surface of the element are known to produce high cross-polarized components off axis, especially in the 45° diagonal plane. Second, the contiguous metallic structure tends to induce parallel plate modes that propagate along the troughs. One way to suppress these modes is to make the elements thicker and to raise the cutoff frequency of the trough region. However, this makes it difficult to interleave to form a dual-pol aperture in the 3D configuration. Third, the tapered slot is not suitable for low profile systems, such as airborne UHF arrays for foliage penetration synthetic aperture radar (FOPEN SAR) and long-range surveillance radars.

For these reasons, a low profile design has been developed. Figure 24-18 shows a flared dipole, also known as a “bunny-ear” element.⁴⁹ This flared dipole is about $1/2\lambda$ long at the high end of the band and $1/8\lambda$ at the low end, supporting a 4:1 band in a large array environment. The low cutoff is made possible by etching out the conducting area near the feed points. This prevents the low-frequency components from being “short-circuited” by the shunt paths between the adjacent elements. Also, the mutual coupling from the adjacent element on each side is about 180° out of phase at midband, and the magnitude is about half of the passive self-reflection term of the S -parameters (S_{11}). The element can be stamped out in a solid plate, or symmetrically printed on both sides of a low- k substrate such as FR4 fiberglass material. It consists of an impedance transformation feed section⁵⁰ and a pair of flared dipole wings. The radiation efficiency of the element is poor when it is electrically small and isolated in free space. However, the efficiency improves significantly in an array environment with the elements lined up in the E -plane.

24.4 WIDEBAND GROUND PLANE

A wideband array becomes band limited once a ground plane is placed behind the radiating aperture. The degradation is caused by the reflection of the back plane, as discussed earlier in the “Characteristics of Wideband Array” subsection. Ideally, the solution is to

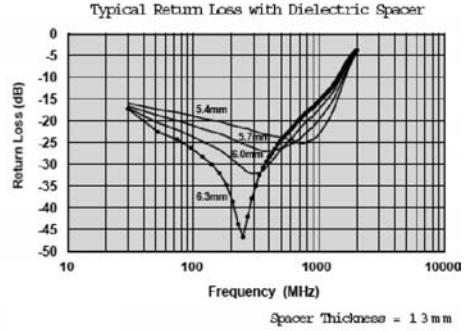
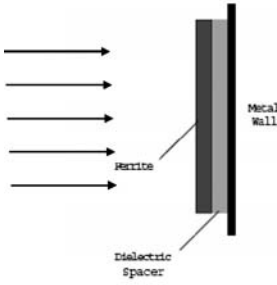


FIGURE 24-19 Ferrite tiles are used in anechoic chamber to absorb power over a wide range of frequency (Courtesy of Fair-Rite Co.).

place a high-Z ground plane⁵¹ behind the aperture to cause total reflection without phase reversal. However, the periodic surface is band-limited due to its resonant nature, so the high-Z ground plane is only useful for narrowband applications so far. A compromise is to use an absorbing material to minimize the reflections from the ground plane, as mentioned in Section 24.3. For certain missions where the bandwidth is more critical than the gain, it may be acceptable to trade off gain, as much as 3 dB, for a wider bandwidth. In this case, the classic Salisbury screen and Jaumann absorber can be used. A comprehensive study of the Salisbury screen can be found in Tuley's chapter of Knott et al's *Radar Cross Section* book,⁵² where extensive discussions on radar absorbing materials (RAM) are given. A general treatment of circuit analog absorbers in the form of multilayer lossy frequency selective surfaces (FSS) is covered in Munk's book.⁵³ Additional references on absorbers are available in the literature.^{54,55} Commercial vendors, such as Emerson and Cuming, also offer multilayer graded absorbers to meet this need. Absorption better than 20 dB over octave band is achievable.

A multilayer resistive absorber normally requires a total thickness of more than $\frac{1}{4}$ wavelength to be effective. For low frequency down to 100-MHz range, this thickness is on the order of 40–80 cm, depending on the dielectric constant of the loading material. Thus alternatives with a lower profile are desirable. One solution is to use ferrite materials. In the construction of a quiet room, a standard approach is to shield the conducting side walls with thin ferrite tiles, as shown in Figure 24-19. The incident impedance of the structure is given by $Z_{in} = j 377 (\mu/\epsilon)^{1/2} \tan(\beta L)$, where β is the wave number in the ferrite material. For a thin slab and low frequency limit, assuming that we are operating in a region where μ is dominated by the imaginary part of the permeability and it has a $1/f$ frequency dependence, the impedance becomes

$$Z_{in} \sim j 377 (\mu/\epsilon)^{1/2} (\beta L) \sim j 377 (2\pi/c) fL (-j\mu''/f) = 377 (2\pi\mu'' L/c) \quad (24-14)$$

By choosing the proper thickness L for a given ferrite material, we can make the factor in the parenthesis equal to unity and the incident impedance equal to 377Ω for low frequency limit. For wide band, a multilayer structure is needed, and it may be feasible to construct a very thin composite absorber using high-index meta-ferrite film material.³⁵ On the other hand, if we could make $\beta L = \pi/2$ by choosing a metamaterial with $\mu\epsilon$ dominated by the real part with $1/f^2$ frequency dependence, then the structure would become a high-Z “ $\lambda/4$ ” reflecting surface for all the frequencies in the band.

However, such metamaterial does not exist.^{56,57} Nevertheless, in the future, multilayer FSS using metallic screens may be replaced by lightweight high index and very thin meta-ferrite film composite materials.⁵⁷

24.5 PROTOTYPE WIDEBAND ARRAYS

Conventional wideband arrays use discrete radiating elements that are usually costly to fabricate for low band applications, and their high profiles are problematic for airborne platforms. To overcome these drawbacks, a long slot array without discrete protruding elements was developed. Figure 24-20a shows a test array of 15×15 elements built by Raytheon for 0.4–2.0 GHz experiment. The lattice spacing of 7.6 cm (3") was chosen to support $\pm 60^\circ$ of grating lobe free scan in both E - and H -planes at the highest frequency. The total thickness of the array between the ground plane and the back plate is 8.1 cm, including a 2-cm layer of air-foam stripline transformer (50- Ω input to 188- Ω “push-pull” balun feedline). The array was only 0.1λ thick at the lowest operating frequency. Additional details are given in Lee et al.³² Another example of the long slot array is shown in Figure 24-20b, which is a compact ultra-lightweight UHF array designed for large-scale airship sensors. The UHF test array has 4×8 (32) elements ($1.12 \text{ m} \times 2.24 \text{ m} \times 23 \text{ cm}$) built and tested for 150–600 MHz operation. The array is 23 cm thick, and it weighs 7 kg (2.8 kg/m^2). Lossy loading material may be used to make it thinner, but this is believed to be the lightest and thinnest UHF array reported for wideband applications without gain penalty. Array patterns, gain, and cross-pol level were measured, and the results were reported in Lee et al.³⁴

Raytheon has built wide-band multifunction arrays capable of simultaneous and time interleaved radar, electronic warfare, and communications functions.⁵⁸ It has also developed a 10:1 bandwidth flared notch (tapered slot) element, operating from 1.8 to 18 GHz, on the DARPA Reconfigurable Aperture (RECAP) Program. The element provides dual orthogonal polarizations with wide scan range capability. A picture of the RECAP element is shown in Figure 24-21. To achieve decade bandwidth, the (cavity-backed) hockey-stick balun and the tapered notch element must be designed to achieve the required bandwidth separately and when cascaded while fitting within the allowed element depth and spacing. The tapered notch was approximated by a large number of slotline segments, each characterized by an

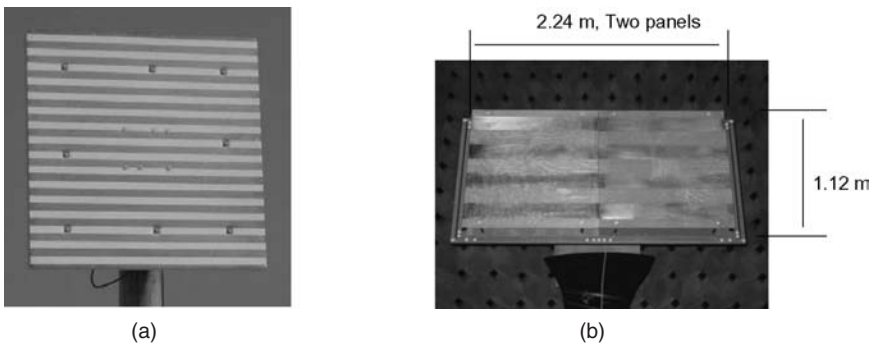


FIGURE 24-20 Wideband long slot arrays: (a) 15×15 elements for 0.4–2.0 GHz band and (b) 4×8 -element UHF array for 150–600 MHz demonstration

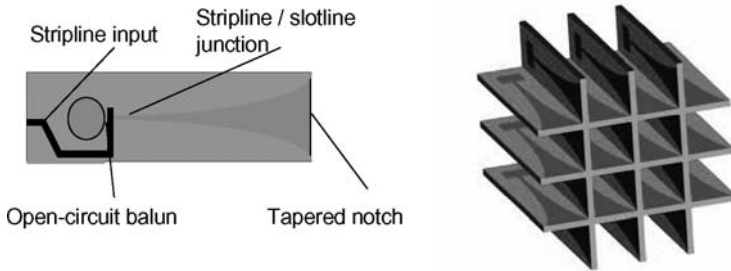
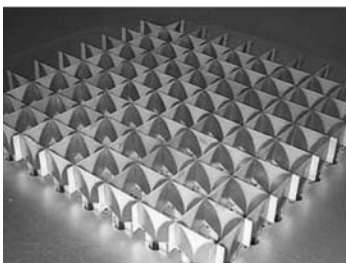


FIGURE 24-21 RECAP Program Dual Polarized Apertures for 1.8–18-GHz operation

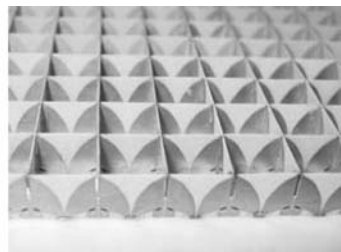
S-matrix. An optimizer was used to determine the tapered notch width profile that provides the minimum element return loss over the specified bandwidth. Note that both the balun and taper separately may not perform satisfactorily at the very low frequency end, but the overall element can be optimized to perform well at the low end limit.

For low profile applications, flared dipole (bunny-ear) elements are more useful. Figure 24-22a shows a 9×9 -element test array designed for low band (1–5 GHz).^{59,60} A partial view of a 25×25 -element high band (3–14 GHz) array is shown in Figure 24-22b. In the experiment,⁶¹ the embedded element pattern was measured, and it tracks with the cosine pattern fairly well up to 60° over the band. The rise time of the impulse response is 100 picoseconds, confirming 10-GHz instantaneous bandwidth. The cross-pol components are 30 dB below at broadside and 20 dB on average up to 60° in all principal planes, a feature not enjoyed by the long tapered slot element.

The flared dipole element was also used in a small UHF antenna for FOPEN SAR application.^{62,63} As shown in Figure 24-23, it is a dual-linearly polarized array consisting of 4×4 elements in the V-pol and 3×4 elements in the H-pol. It operates over a 4:1 bandwidth (150–600 MHz), with VSWR less than 2:1 over the band. The “egg-crate” array was designed to fit in a small radome (50 cm \times 100 cm \times 100 cm). Each set of radiating elements produces a side-looking beam, shaped to cover a broad region in the elevation and azimuth plane, with the beam peak pointed in the direction of 30° depression angle from the horizon.



(a)



(b)

FIGURE 24-22 Wideband flared dipole (bunny-ear) test arrays: (a) 9×9 elements for 1.0–5.0 GHz band and (b) 25×25 -element X-band array for 3–14 GHz demonstration

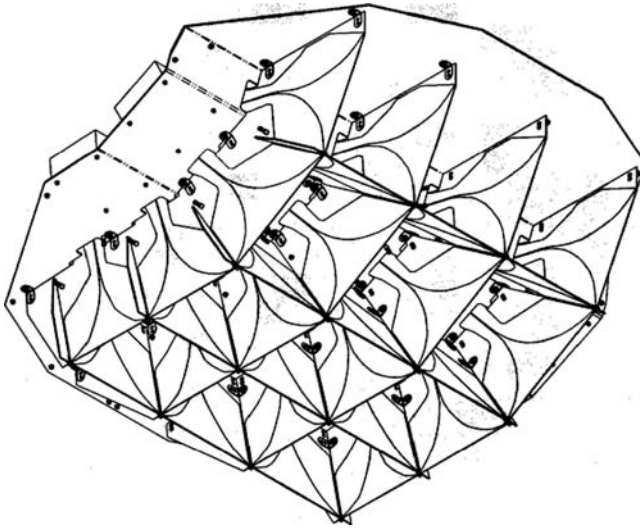


FIGURE 24-23 UHF antenna with flared dipole elements for FOPEN SAR application

REFERENCES

1. S. Silver, *Microwave Antenna Theory and Design* (New York: McGraw-Hill, 1949).
2. A. W. Love, *Electromagnetic Horn Antennas* (New York: IEEE Press, 1976).
3. R. S. Elliott, *Antenna Theory and Design* (Englewood Cliffs, NJ: Prentice Hall, 1981).
4. A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, (eds.), *The Handbook of Antenna Design* (London: IEE/Peter Peregrinus, 1983).
5. Y. Y. Lo and S. W. Lee, *Antenna Handbook* (New York: Van Nostrand Reinhold, 1988).
6. J. D. Kraus, *Antennas*, 2nd Ed. (New York: McGraw-Hill, 1988).
7. Y. T. Lo and S. W. Lee, *Antenna Handbook*, vols. I–III (New York: Van Nostrand Reinhold, 1993).
8. R. C. Johnson and H. Jasik (eds.), *Antenna Engineering Handbook* (New York: McGraw-Hill, 1993).
9. D. Pozar and D. Schaubert, *Microstrip Antennas* (Piscataway, NJ: IEEE Press, 1995).
10. C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd Ed. (New York: John Wiley & Sons, Inc., 1996).
11. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd Ed. (New York: John Wiley & Sons, Inc., 1998).
12. R. C. Hansen, *Microwave Scanning Antennas* (New York: Academic Press, 1964) and (Tallahassee: Peninsula Publishing, 1986).
13. A. A. Oliner and G. H. Knittel, *Phased Array Antennas* (Dedham, MA: Artech House, 1972).
14. N. Amitay, V. Galindo, and C. P. Wu, *Theory and Analysis of Phased Array Antennas* (New York: Wiley-Interscience, 1972).
15. M. T. Ma, *Theory and Application of Antenna Arrays* (New York: Wiley-Interscience, 1974).
16. R. J. Mailloux, *Phased Array Antenna Handbook* (Dedham, MA: Artech House, 1994).

17. R. C. Hansen, *Phased Array Antennas* (New York: Wiley-Interscience, 1998).
18. H. A. Wheeler, "The Radiation Resistance of an Antenna in an Infinite Array or Waveguide," *Proc. IRE*, vol. 36 (April 1948): 478–488.
19. H. A. Wheeler, "Simple Relations Derived from a Phased-Array Antenna Made of an Infinite Current Sheet," *IEEE Trans. Antennas Propagat.* (1965): 506.
20. R. C. Hansen, *Phased Array Antennas*, Chap. 7 (New York: Wiley-Interscience, 1998).
21. R. C. Hansen, "Current Induced on a Wire: Implications for Connected Arrays," *IEEE Antennas and Wireless Propagat. Letters*, vol. 2 (2003): 288.
22. R. C. Hansen, "Linear Connected Arrays," *IEEE Antennas and Wireless Propagat. Letters*, vol. 3 (2004): 154–156.
23. R. C. Hansen, "Dipole Arrays with Non-Foster Circuits," *Proc. IEEE International Symp. on Phased Array Systems and Technology* (2003): 40.
24. R. C. Hansen, "Non-Foster and Connected Planar Arrays," *Radio Science*, vol. 39, RS4004 (2004): 1–14.
25. A. Neto and J. J. Lee, "'Infinite Bandwidth' Long Slot Array Antenna," *IEEE Trans. Antennas Propagat.*, vol. 4 (2005): 75.
26. R. C. Hansen, "Comments on 'Infinite Bandwidth' Long Slot Array Antenna," *IEEE Antennas Wireless Propagat. Letters*, vol. 4 (2005): 497.
27. A. Neto and J. J. Lee, "Ultrawide-Band Properties of Long Slot Arrays," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 2 (2006): 534–543.
28. R. S. Elliott, *Antenna Theory and Design*, Chap. 7 (Englewood Cliffs, NJ: Prentice Hall, 1981): 336–341.
29. H. G. Booker, "Slot Aerials and Their Relation to Complementary Wire Aerials (Babinet's Principle)," *J. IEE*, 93, pt. IIIA (1946): 620–626.
30. R. S. Elliott, *Antenna Theory and Design*, Chap. 7 (Englewood Cliffs, NJ: Prentice Hall, 1981): 278–284.
31. J. J. Lee, "Effects of Metal Fences on the Scan Performance of an Infinite Dipole Array," *IEEE Trans. Antennas Propagat.*, vol. 38 (1990): 683–692.
32. J. J. Lee, S. Livingston, R. Koenig, D. Nagata and L. Lai, "Compact Light Weight UHF Arrays Using Long Slot Apertures," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 2 (2006): 2009–2015.
33. J. J. Lee, S. Livingstone and R. Koenig, "Wide Band Slot Array Antennas," *Proc. AP-S Symp.*, vol. 2 (June 2003): 452–455.
34. J. J. Lee, S. Livingston, R. Koenig, D. Nagata, L. Lai and A. Neto, "Long Slot Arrays - Part 2: Ultra Wideband Test Results," *Proc. IEEE A.P. Symp.*, July 2005.
35. For example, meta-ferrite composites developed by R. Walser, Univ. Texas, Austin, and Meta-Materials, Inc., 2225 W. Braker Lane, Austin, TX 78758.
36. C. E. Baum, "Some Characteristics of Planar Distributed Sources for Radiating Transient Pulses," *Sensor and Simulation Note #100*, AFRL (March 1970).
37. C. E. Baum, *Transient Arrays, Ultra-Wideband, Short-Pulse Electromagnetics 3*, C. Baum, L. Carin and A. Stone, (eds.) (New York: Plenum Press, 1997): 129–138.
38. D. T. McGrath and C. E. Baum, "Scanning and Impedance Properties of TEM Horn Arrays for Transient Radiation," *IEEE Trans. Antennas Propagat.*, vol. 47 (March 1999): 469–473.
39. N. Inagaki, Y. Isogai and Y. Mushiaki, "Ichimatsu Moyou Antenna—Self-Complementary Antenna with Periodic Feeding Points," *IECE Transactions*, vol. 62, (1979): 388–395.
40. M. P. Kesler, J. G. Maloney, and G. S. Smith, "FDTD Analysis of Novel Antenna Array and Substrate Concepts," ICAP/JINA Conf. on AP—Davos, Switzerland, April 2000.
41. P. Friederich et al., "A New Class of Broadband Planar Apertures," *Proc. Allerton Antenna Applications Symp.* (September 2001): 561–587.
42. B. Munk, *Finite Antenna Arrays and FSS*, Chap. 6 (New York: John Wiley & Sons Inc, 2003).
43. A. Q. Boryszenko and D. H. Schaubert, "Physical Aspects of Mutual Coupling in Finite Broadband Tapered Slot (Vivaldi) Arrays," 2005 5th International Conference on Antenna Theory and Techniques, Kiev, Ukraine, May 24–27.

44. C. Craeye, A. G. Tijhuis and D. H. Schaubert, "An Efficient MoM Formulation for Finite-by-Infinite Arrays of Two-Dimensional Antennas Arranged in a Three-Dimensional Structure," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 1 (2004): 271.
45. D. H. Schaubert, T. -H. Chio and H. Holter, "TSA Element Design for 500–1500 MHz Array," *IEEE Proc. AP-S Symp.*, vol. 1 (2000): 178–181.
46. H. Holter, T. -H. Chio and D. H. Schaubert, "Experimental Results of 144-Element Dual-Polarized Endfire Tapered-Slot Phased Arrays," *IEEE Trans. Antennas Propagat.*, vol. 48, no. 11 (2000): 1707–1718.
47. R. N. Simons and R. Q. Lee, "Linearly Tapered Slot Antenna Impedance Characteristics," *IEEE Proc. AP-S Symp.* (1995): 170–173.
48. R. N. Simons and R. Q. Lee, "Impedance Matching of Tapered Slot Antenna Using a Dielectric Transformer," *Electronics Letters*, vol. 34, no. 24 (1998): 2287–2289.
49. J. J. Lee and S. Livingston, "Wideband Bunny-Ear Radiating Element," *IEEE Antennas & Propagat. Symp. Proc.* (1993): 1604.
50. J. J. Lee, "Slotline Impedance," *IEEE Trans. Microwave Theory and Tech.*, vol. 39 (April 1991): 666–672.
51. D. Sievenpiper, L. Zhang, R. Broas, N. Alexopolous, and E. Yablonovitch, "High-Impedance Electromagnetic Surfaces with a Forbidden Frequency Band," *IEEE Trans. Microwave Theory Tech.*, vol. 47 (November 1999): 2059–2074.
52. E. F. Knott et al, *Radar Cross Section*, Chap. 9, (Dedham, MA: Artech House, 1985).
53. B. Munk, *Frequency Selective Surfaces*, Chap. 9 (New York: John Wiley & Sons, 2000).
54. E. F. Knott, "The Thickness Criterion for Single-Layer Radar Absorbents," *IEEE Trans. Antennas Propagat.*, vol. 27, no. 5 (1979): 698.
55. E. F. Knott and C. D. Lunden, "The Two-Sheet Capacitive Jaumann Absorber," *IEEE Trans. Antennas Propagat.*, vol. 43, no. 11 (1995): 1339.
56. J. L. Wallace, "Broadband Magnetic Microwave Absorbers: Fundamental Limitations," *IEEE Trans. on Magnetics*, vol. 29, no. 6 (1993): 4209.
57. Private communication with Prof. Rodger Walser, Univ. of Texas, Austin, and Meta-Materials, Inc., 2225 W. Braker Lane, Austin, TX 78758.
58. C. Hemmi, R. T. Dover, F. German, and A. Vespa, "Multifunction Wide-Band Array Design," *IEEE Trans. Antennas Propagat.*, vol. 47, no. 3 (1999): 425–431.
59. J. J. Lee, S. Livingston, et al "A Low-Profile Wide Band (5:1) Dual-Pol Array," *IEEE Antenna and Wireless Propagation Letters*, vol. 2 (2003): 46.
60. J. J. Lee, S. Livingston, et al, Correction to "A Low-Profile Wide-Band (5:1) Dual-Pol Array," *IEEE Antennas and Wireless Propagation Letters*, vol. 3 (2004): 375.
61. J. J. Lee, S. Livingston and R. Koenig, "Performance of a Wideband (3–14 GHz) Dual-Pol Array," *IEEE Antennas and Propagation Society International Symp.*, (2004): 551.
62. J. J. Lee, S. Livingston, et al, "A UHF Wide-Band SAR Antenna," *Proc. of IEEE Int'l Conference on Phased Array Systems and Technology* (2000): 437.
63. J. J. Lee, S. Livingston, et al, "Performance of an Ultra Wide Band UHF Antenna," *IEEE Antennas & Propagat. Symp. Proc.* (2000): 1198.

Chapter 25

Smart Antennas

Frank B. Gross

Argon ST, Fairfax, Virginia

CONTENTS

25.1	INTRODUCTION.....	25-2
25.2	THE HISTORICAL DEVELOPMENT OF SMART ANTENNAS	25-2
25.3	ADAPTIVE BEAMFORMING	25-3

25.1 INTRODUCTION

Modern *beamsteered* arrays, where the pattern is shaped according to various optimum criteria, are called smart antennas. Smart antennas have alternatively been referred to as *digital beamformed* (DBF) arrays or *adaptive* arrays (when adaptive algorithms are employed). The term “smart” implies the use of signal processing in order to shape the beam pattern according to certain conditions. Smart antennas hold the promise for improved sensing systems,¹⁻⁶ improved system capacities with mobile wireless, and improved wireless communications⁷⁻⁹ through the implementation of space division multiple access (SDMA).

Smart antenna patterns are controlled by algorithms based upon optimizing certain criteria. These criteria are typically based on maximizing the signal-to-interference ratio (SIR), minimizing the variance, minimizing the mean-square error (MSE), steering toward a signal of interest, or nulling depth against interfering signals. These criteria can be used in static fixed geometries or in dynamic tracking of moving emitters and interference.

Although there are many smart antenna algorithms, we will only review the canonical methods that include least mean squares (LMS), sample matrix inversion (SMI), recursive least squares (RLS), conjugate gradient method (CGM), constant modulus algorithm (CMA), and the least squares constant modulus algorithm (LS-CMA).

25.2 THE HISTORICAL DEVELOPMENT OF SMART ANTENNAS

The development of smart antennas began in the late 1950s. The term “adaptive array” was first coined by Van Atta¹⁰ in 1959 to describe a *self-phased* array. Self-phased arrays merely reflected all incident signals back in the direction of arrival by using phase conjugation schemes. Self-phasing has alternatively been called *retrodirection*. Self-phased arrays are instantaneously adaptive arrays since they essentially reflect the incident signal in a fashion similar to the classic corner reflector.

Phase-locked loop (PLL) systems were incorporated into arrays in the 1960s in an effort to construct better retrodirective arrays.¹¹ PLLs are still used in single beam scanning systems.¹²

Adaptive sidelobe cancellation (SLC) was first proposed by Howells^{13,14} in 1959. This technique allowed for interference nulling, thus raising the signal-to-interference ratio. The Howells SLC was the first adaptive scheme that allowed for automatic interference nulling. By maximizing the generalized signal-to-noise ratio, Applebaum developed the algorithm governing adaptive interference cancellation,^{15,16} which is now referred to as the Howells-Applebaum algorithm. At the same time, through the use of least mean squares (LMS), Widrow and others applied self-training to adaptive arrays.^{17,18} The Howells-Applebaum and Widrow algorithms are both steepest-descent/gradient-search methods that converge to the optimum Wiener solution. The convergence rate of these methods is dependent upon the eigenvalue spread¹⁹ such that larger spreads require longer convergence times. The convergence time constant is given by Monzingo and Miller²⁰:

$$\tau_i = \frac{1}{2\mu\lambda_i} \quad (25-1)$$

where

μ = gradient step size

λ_i = i th eigenvalue

The eigenvalues are derived from the covariance matrix of array voltages. The largest eigenvalues correspond to the strongest signals.

Since the convergence of the SLC algorithm is slow for large eigenvalue spreads, Reed, Mallett, and Brennen²¹ developed the direct sample matrix inversion (SMI) technique in 1974. Another means of improving the eigenvalue limited convergence rate is through the use of conjugate gradient methods. Most notably, Powell's method²² has been applied to steepest descent problems.

In many modern communication systems, the signals of interest are phase or frequency modulated. Thus the signal amplitude, or modulus, should be constant in the absence of multipath and fading. As a consequence of the constant modulus property, many popular blind adaptive algorithms have been developed. These include the Godard constant modulus algorithm (CMA) proposed in 1980²³ and the least squares constant modulus algorithm (LS-CMA) proposed by Agee²⁴ in 1986.

A complementary advance in smart antennas came with the application of spectral estimation methods to array processing. In 1969 Capon²⁵ used a maximum likelihood (ML) method to solve for the minimum variance distortionless response (MVDR) of an array. In 1972 Burg applied the maximum entropy method to spectral estimation, and his technique was soon adapted to array signal processing.^{26,27} In 1973 Pisarenko²⁸ developed the harmonic decomposition technique based upon minimizing the mean-squared error under the constraint that the norm of the weight vector be equal to unity. The minimum-norm (min-norm) method was developed by Reddi²⁹ in 1979 and improved by Kumaresan and Tufts³⁰ in 1983. The min-norm algorithm optimizes the weight vector by solving the optimization problem where the weight vector is orthogonal to the signal eigenvector subspace. The now famous MUSIC algorithm developed by Schmidt³¹ in 1986 is a spectral estimation algorithm that exploits the orthogonality of the noise subspace of the array covariance matrix. The ESPRIT technique was first proposed by Roy and Kailath³² in 1989. The goal of ESPRIT is to exploit the rotational invariance in the signal subspace that is created by two arrays with a translational invariance structure.

Many adaptive methods allow the user to achieve higher resolutions than that permitted by the beamwidth of the array. The price of increased resolution comes at the cost of greater computational intensity.

25.3 ADAPTIVE BEAMFORMING

The fixed beamforming approaches assume a stationary geometry with a fixed emitter angle of arrival (AOA). If the signal AOAs do not change with time, the optimum array weights do not need to be updated after they are initially found. However, if the desired arrival angles change with time, it is necessary to devise an optimization scheme that operates "on-the-fly" to continually recalculate the optimum array weights. The receiver signal processing algorithm must allow for a stepwise update of weights in order to adapt to the continuously changing electromagnetic environment. The adaptation process must satisfy a specified optimization criterion. Several popular optimization techniques include least mean squares (LMS), sample matrix inversion (SMI), recursive least squares (RLS), the conjugate gradient method (CGM), and the constant modulus algorithm (CMA). Each of these techniques is described in the sections that follow.

Least Mean Squares

The least mean squares algorithm is a gradient-based approach. Monzingo and Miller²⁰ give an excellent fundamental treatment of this approach. Gradient-based algorithms assume an

established quadratic performance surface. The performance surface (or cost function), $C(\bar{w})$, is in the shape of an elliptic paraboloid having one minimum. One of the best ways to establish the minimum is through the use of a gradient method by minimizing the cost function. Consider the adaptive array shown in Figure 25-1.

The M -element array has M weights that can be adjusted to modify the beam pattern. It is generally assumed that there is one arriving signal of interest $s(k)$ with N arriving interferers. One time sample is represented by k . The m th array element has an output, $x_m(k)$, which consists of the desired signal, interfering signals, and noise. Often a desired signal $d(k)$ is employed that is representative of the signal $s(k)$ and that can be used to “train” the array weights in order to enhance the signal $s(k)$ at the array output. An error signal, $\varepsilon(k)$, is generally calculated as the difference between the array output, $y(k)$, and the desired signal, $d(k)$. The error can be minimized by minimizing the mean squared error (MSE).

The error, as shown in Figure 25-1, is given as

$$\varepsilon(k) = d(k) - \bar{w}^H(k)\bar{x}(k) \quad (25-2)$$

The squared error is given as

$$|\varepsilon(k)|^2 = |d(k) - \bar{w}^H(k)\bar{x}(k)|^2 \quad (25-3)$$

Through some simple algebra, it can be shown that the mean-square error is given by

$$|\varepsilon(k)|^2 = |d(k)|^2 - 2d(k)\bar{w}^H\bar{x}(k) + \bar{w}^H\bar{x}(k)\bar{x}^H(k)\bar{w} \quad (25-4)$$

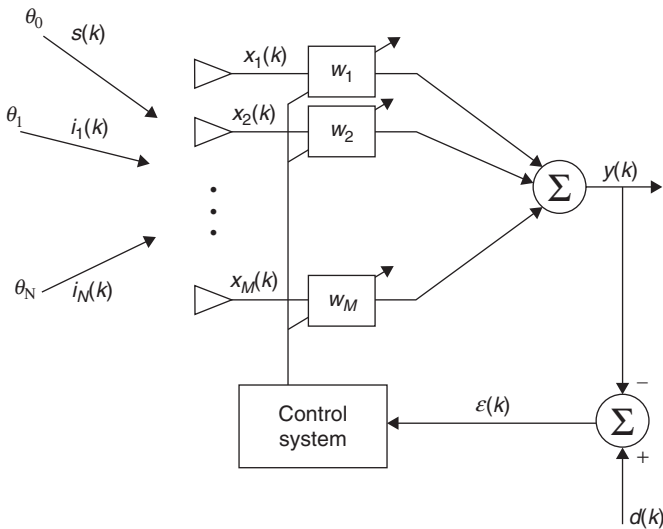


FIGURE 25-1 Typical adaptive array block diagram

For the purposes of simplification, we will suppress the time dependence notation k . Taking the expected value of both sides and simplifying the expression, we get

$$E[|\varepsilon|^2] = E[|d|^2] - 2\bar{w}^H \bar{r} + \bar{w}^H \bar{R}_{xx} \bar{w} \quad (25-5)$$

where the following vector and matrix covariances are defined:

$$\bar{r} = E[d^* \cdot \bar{x}] = E[d^* \cdot (\bar{x}_s + \bar{x}_i + \bar{n})] \quad (25-6)$$

$$\bar{R}_{xx} = E[\bar{x}\bar{x}^H] = \bar{R}_{ss} + \bar{R}_{nn} \quad (25-7)$$

$$\bar{R}_{ss} = E[\bar{x}_s \bar{x}_s^H] \quad (25-8)$$

$$\bar{R}_{nn} = \bar{R}_{ii} + \bar{R}_{nn} \quad (25-9)$$

The expression in Eq. 25-5 is a quadratic function of the weight vector \bar{w} . This function is sometimes called the “performance surface” or “cost function” and forms a quadratic surface in M -dimensional space. Since the optimum weights provide the minimum MSE, the extremum is the minimum for this function. A trivial example of a performance surface is given for a two-element array that produces a two-dimensional surface as depicted in Figure 25-2. As the desired angle-of-arrival changes with time, the quadratic surface minimum changes with time in the weight plane.

The cost function of Eq. 25-5 can be rewritten as

$$C(\bar{w}) = D - 2\bar{w}^H \bar{r} + \bar{w}^H \bar{R}_{xx} \bar{w} \quad (25-10)$$

where

$$D = E[|d|^2]$$

The performance surface is determined from the gradient of Eq. 25-10 with respect to the weight vector. Thus

$$\nabla_{\bar{w}}(C(\bar{w})) = 2\bar{R}_{xx} \bar{w} - 2\bar{r} \quad (25-11)$$

The minimum occurs when the gradient is zero. With the left side of Eq. 25-11 set to zero, the optimum Wiener solution for array weights is given by

$$\bar{w}_{\text{opt}} = \bar{R}_{xx}^{-1} \bar{r} \quad (25-12)$$

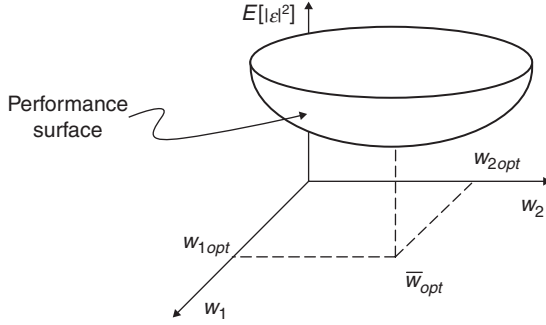


FIGURE 25-2 Quadratic surface for MSE

The Wiener solution of Eq. 25-12 is based on assuming knowledge of all signal statistics. In general, we do not know the signal statistics and must resort to estimating both the array covariance matrix (\hat{R}_{xx}) and the signal covariance vector (\hat{r}) at an instant in time or over a block of time samples. The instantaneous estimates of array covariance and (ideal) signal covariance are given as

$$\hat{R}_{xx}(k) \approx \bar{x}(k)\bar{x}^H(k) \quad (25-13)$$

and

$$\hat{r}(k) \approx d^*(k)\bar{x}(k) \quad (25-14)$$

We can employ an iterative technique called the *method of steepest descent* to approximate the gradient of the cost function. The method of steepest descent can be approximated in terms of the weights using the least mean squares (LMS) method advocated by Widrow et al.^{17,18} The direction of steepest descent is in the opposite direction of the gradient vector and, in an iterative approximation, is given as

$$\bar{w}(k+1) = \bar{w}(k) - \frac{1}{2} \mu \nabla_{\bar{w}}(C(\bar{w}(k))) \quad (25-15)$$

where

μ = step-size parameter

$\nabla_{\bar{w}}$ = gradient of the performance surface

If we substitute the instantaneous covariance approximations of Eqs. 25-13 and 25-14 and the performance surface gradient from Eq. 25-11, the LMS solution can now be written as

$$\begin{aligned} \bar{w}(k+1) &= \bar{w}(k) - \mu [\hat{R}_{xx} \bar{w} - \hat{r}] \\ &= \bar{w}(k) + \mu e^*(k) \bar{x}(k) \end{aligned} \quad (25-16)$$

where

$$e(k) = d(k) - \bar{w}^H(k) \bar{x}(k) = \text{error signal}$$

The convergence of the LMS algorithm is directly proportional to the step-size parameter μ . If the step-size is too small, the convergence is slow (overdamped case). If the convergence is slower than the changing angles of arrival, it is possible that the adaptive array cannot acquire the signal of interest and therefore cannot track the changing signal. If the step-size is too large, the LMS algorithm will overshoot the optimum weights of interest (underdamped case). Under this condition, convergence is too fast, and the weights will oscillate about the optimum weights, but will not accurately track the desired solution. It is therefore imperative to choose a step-size in a range that ensures optimal convergence. Stability of convergence is provided when the following condition is met:²⁰

$$0 \leq \mu \leq \frac{1}{2\lambda_{\max}} \quad (25-17)$$

where

$$\lambda_{\max} = \text{the largest eigenvalue of } \hat{R}_{xx}$$

Since the covariance matrix is positive definite, all eigenvalues are positive. If all the interfering signals are noise and there is only one signal of interest, we can approximate the condition in Eq. 25-17 as

$$0 \leq \mu \leq \frac{1}{2 \text{ trace}[\hat{R}_{xx}]} \quad (25-18)$$

As an example of the LMS routine, let us model an $M = 8$ element array with uniform element spacing of $\lambda/2$. Assume the desired signal is arriving at the angle $\theta_0 = 30^\circ$ and an interferer arrives at the angle $\theta_1 = -60^\circ$. Allowing the step size to be $\mu = .02$, Figure 25-3

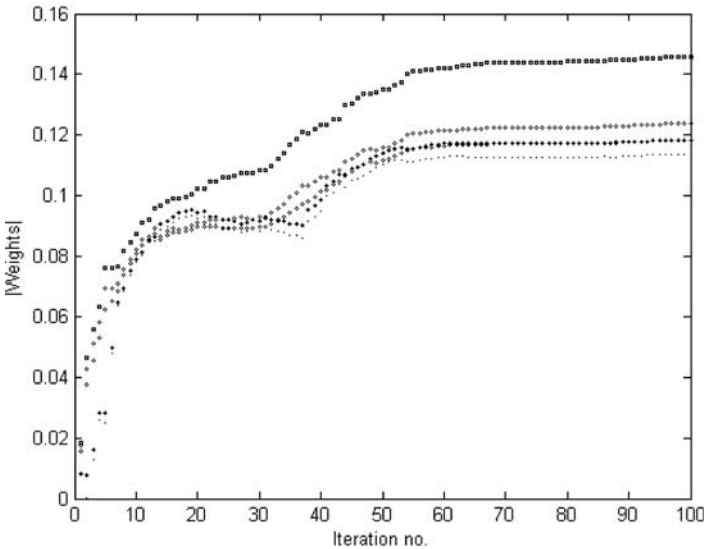


FIGURE 25-3 Magnitude of array weights

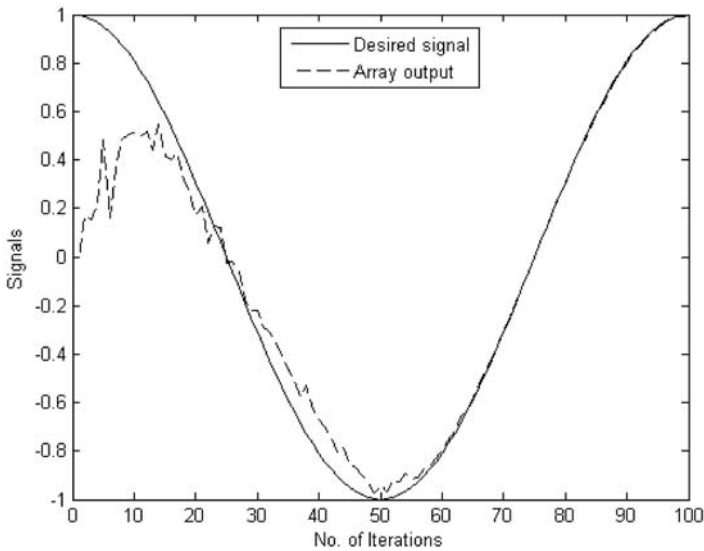


FIGURE 25-4 Acquisition and tracking of desired signal

shows the resulting magnitude of antenna element weights over 100 iterations. Figure 25-4 shows how the array output acquires and tracks the desired signal over 100 iterations. Figure 25-5 shows how the resulting mean square error converges to nearly zero after 25 iterations. Figure 25-6 shows the final weighted array pattern, which as expected, has a peak at the desired direction of 30° and a null at the interfering direction of -60° .

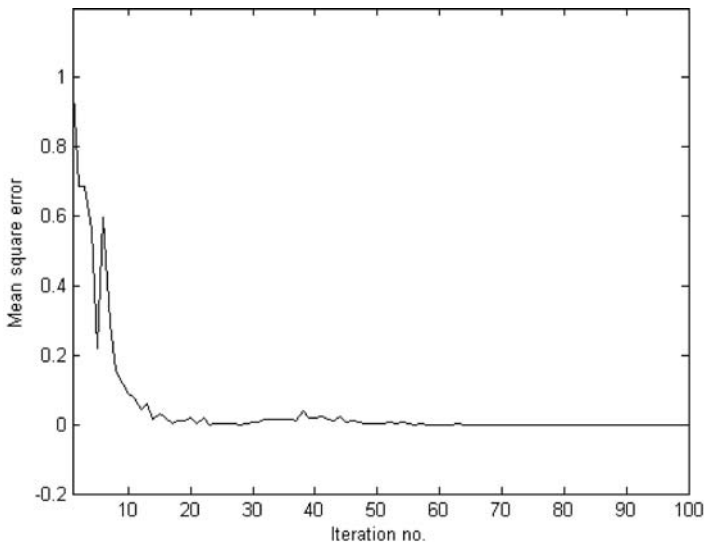


FIGURE 25-5 Mean square error

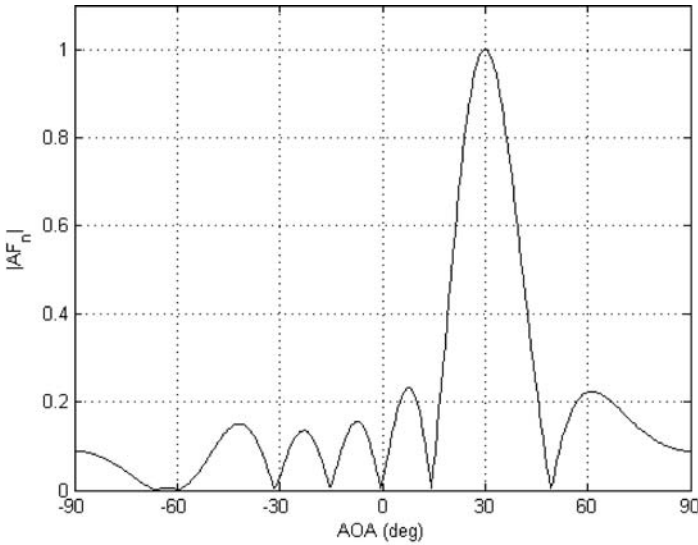


FIGURE 25-6 Weighted LMS array

Sample Matrix Inversion

One of the drawbacks of the LMS adaptive scheme is that the algorithm must go through many iterations before satisfactory convergence is achieved. If the signal characteristics are rapidly changing, the LMS adaptive algorithm may not allow tracking of the desired signal in a satisfactory manner. The rate of convergence of the weights is dictated by the eigenvalue spread of the array covariance matrix, as indicated by Eq. 25-1. In the previous example, the LMS algorithm did not converge until after 25 iterations. One possible approach to circumventing the relatively slow convergence of the LMS scheme is by use of sample matrix inversion (SMI).^{7,26,33} This method is also known as direct matrix inversion (DMI).²⁰ The “sample matrix” is a time average estimate of the array covariance matrix using K time samples. If the random process is ergodic, then the time average estimate of covariance matrix will rapidly approach the actual estimate of the covariance matrix.

Recalling the earlier discussion of the minimum mean square error (MSE), the optimum array weights are given by the optimum Wiener solution as

$$\bar{\mathbf{w}}_{\text{opt}} = \bar{\mathbf{R}}_{xx}^{-1} \bar{\mathbf{r}} \quad (25-19)$$

where

$$\bar{\mathbf{r}} = E[d^* \cdot \bar{\mathbf{x}}]$$

$$\bar{\mathbf{R}}_{xx} = E[\bar{\mathbf{x}} \bar{\mathbf{x}}^H]$$

Under conditions of covariance ergodicity, we can estimate the covariance matrix by calculating the time average of a block of data such that

$$\hat{R}_{xx} = \frac{1}{K} \sum_{k=1}^K \bar{x}(k) \bar{x}^H(k) \quad (25-20)$$

where

K = number of data samples in a block

In similar fashion, the covariance vector \bar{r} can be estimated by the time average

$$\hat{r} = \frac{1}{K} \sum_{k=1}^K d^*(k) \bar{x}(k) \quad (25-21)$$

Since we use a K -length block of data, this method is called a *block-adaptive* approach. Thus the algorithm adapts the weights block-by-block.

Define the matrix $\bar{X}_K(k)$ as the k th block of \bar{x} column vectors ranging over K data snapshots. (Note the switch in use of k from Eq. 25-21 to the use of k representing block number in the following equations. k still represents an iteration step.) Thus

$$\bar{X}_K(k) = \begin{bmatrix} x_1(1+kK) & x_1(2+kK) & \cdots & x_1(K+kK) \\ x_2(1+kK) & x_2(2+kK) & & \vdots \\ \vdots & & \ddots & \\ x_M(1+kK) & \cdots & & x_M(K+kK) \end{bmatrix} \quad (25-22)$$

where

k = block number

K = block length

Thus the estimate of the array covariance matrix is given by

$$\hat{R}_{xx}(k) = \frac{1}{K} \bar{X}_K(k) \bar{X}_K^H(k) \quad (25-23)$$

In addition, the desired signal vector can be defined by

$$\bar{d}(k) = [d(1+kK) \quad d(2+kK) \quad \cdots \quad d(K+kK)] \quad (25-24)$$

Thus the estimate of the covariance vector is given by

$$\hat{r}(k) = \frac{1}{K} \bar{d}^*(k) \bar{X}_K(k) \quad (25-25)$$

The SMI weights can then be calculated for the k th block of length K as

$$\begin{aligned} \bar{w}_{\text{SMI}}(k) &= \bar{R}_{xx}^{-1}(k) \bar{r}(k) \\ &= [\bar{X}_K(k) \bar{X}_K^H(k)]^{-1} \bar{d}^*(k) \bar{X}_K(k) \end{aligned} \quad (25-26)$$

As an example of the SMI technique, again we use an $M = 8$ element array with uniform spacing $\lambda/2$. Allow the desired signal to arrive at the angle $\theta_0 = 30^\circ$ and an interferer to arrive at $\theta_1 = -60^\circ$. Let the block length be $K = 30$. Assume the noise variance is $\sigma_n^2 = .01$.

The optimum Wiener weights are found, and the resulting array pattern is seen in Figure 25-7. The SMI pattern is similar to the LMS pattern and was generated with no iterations, albeit at the cost of collecting K data samples.

The SMI algorithm, although faster than the LMS algorithm, has several drawbacks. The covariance matrix may be ill conditioned, resulting in errors or singularities when inverted. In addition, for large arrays, there is the challenge of inverting large matrices. To invert the covariance matrix of size $M \times M$ requires $M^3/2 + M^2$ complex multiplications.²¹ The SMI update frequency will necessarily depend on signal frequencies and on channel fading conditions.

Recursive Least Squares

Even though the SMI method is faster than the LMS algorithm, the computational burden and potential singularities can cause problems. However, we can recursively calculate the required covariance matrix and the required covariance vector in a method called recursive least squares (RLS). Recall that in Eqs. 25-23 and 25-25 estimates of the covariance matrix and vector were taken as the sum of the terms divided by the block length K . When we calculate the weights in Eq. 25-26, the division by K is cancelled. Thus we can simplify the expressions for the covariance matrix and the covariance vector:

$$\hat{R}_{xx}(k) = \sum_{i=1}^k \bar{x}(i)\bar{x}^H(i) \quad (25-27)$$

$$\bar{r}(k) = \sum_{i=1}^k d^*(i)\bar{x}(i) \quad (25-28)$$

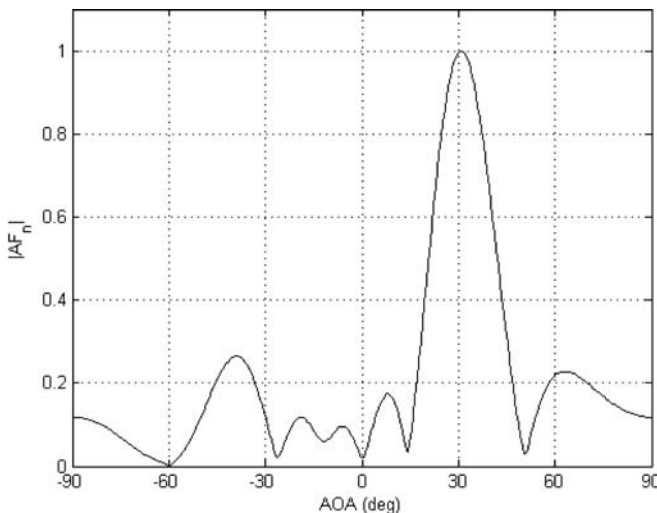


FIGURE 25-7 Weighted SMI array pattern

where

k = block length and last time sample k

$\hat{R}_{xx}(k), \bar{r}(k)$ = covariance estimates ending at time sample k

Both of the preceding summations use rectangular windows, and thus they equally consider all previous time samples. Since the signal sources can change or slowly move with time, we might want to deemphasize the earliest data samples and emphasize the most recent ones. This can be accomplished by modifying Eqs. 25-27 and 25-28 such that we forget the earliest time samples. This is called a *weighted estimate*. Thus

$$\hat{R}_{xx}(k) = \sum_{i=1}^k \alpha^{k-i} \bar{x}(i) \bar{x}^H(i) \quad (25-29)$$

$$\hat{r}(k) = \sum_{i=1}^k \alpha^{k-i} d^*(i) \bar{x}(i) \quad (25-30)$$

where

α = “forgetting” factor

The “forgetting” factor is also sometimes referred to as the exponential weighting factor.³⁴ α is a positive constant such that $0 \leq \alpha \leq 1$. When $\alpha = 1$ we restore the ordinary least squares algorithm. $\alpha = 1$ also indicates infinite memory. Let us break up the summation in Eqs. 25-29 and 25-30 into two terms: the summation for values up to $i = k - 1$ and the last term for $i = k$.

$$\begin{aligned} \hat{R}_{xx}(k) &= \alpha \sum_{i=1}^{k-1} \alpha^{k-1-i} \bar{x}(i) \bar{x}^H(i) + \bar{x}(k) \bar{x}^H(k) \\ &= \alpha \hat{R}_{xx}(k-1) + \bar{x}(k) \bar{x}^H(k) \end{aligned} \quad (25-31)$$

$$\begin{aligned} \hat{r}(k) &= \alpha \sum_{i=1}^{k-1} \alpha^{k-1-i} d^*(i) \bar{x}(i) + d^*(k) \bar{x}(k) \\ &= \alpha \hat{r}(k-1) + d^*(k) \bar{x}(k) \end{aligned} \quad (25-32)$$

Thus future values for the array covariance estimate and the vector covariance estimate can be found using previous values. The advantage of the recursion approach is that you need not calculate the covariance for an entire block of length k . Rather, each update only requires the current time sample and the previous covariance matrix.

Not only can we recursively calculate the most recent covariance estimates, we can also use Eq. 25-31 to derive a recursion relationship for the inverse of the covariance matrix. The next steps follow the derivation in Haykin.³⁴ We can invoke the Sherman Morrison-Woodbury (SMW) theorem³⁵ to find the inverse of Eq. 25-31. The SMW theorem states

$$(\bar{A} + \bar{z} \bar{z}^H)^{-1} = \bar{A}^{-1} - \frac{\bar{A}^{-1} \bar{z} \bar{z}^H \bar{A}^{-1}}{1 + \bar{z}^H \bar{A}^{-1} \bar{z}} \quad (25-33)$$

Applying Eq. 25-33 to Eq. 25-31, we have the following recursion formula:

$$\hat{R}_{xx}^{-1}(k) = \alpha^{-1} \hat{R}_{xx}^{-1}(k-1) - \frac{\alpha^{-2} \hat{R}_{xx}^{-1}(k-1) \bar{x}(k) \bar{x}^H(k) \hat{R}_{xx}^{-1}(k-1)}{1 + \alpha^{-1} \bar{x}^H(k) \hat{R}_{xx}^{-1}(k-1) \bar{x}(k)} \quad (25-34)$$

We can simplify Eq. 25-34 by defining the gain vector $\bar{g}(k)$

$$\bar{g}(k) = \frac{\alpha^{-1} \hat{R}_{xx}^{-1}(k-1) \bar{x}(k)}{1 + \alpha^{-1} \bar{x}^H(k) \hat{R}_{xx}^{-1}(k-1) \bar{x}(k)} \quad (25-35)$$

that transforms Eq. 25-34 into

$$\hat{R}_{xx}^{-1}(k) = \alpha^{-1} \hat{R}_{xx}^{-1}(k-1) - \alpha^{-1} \bar{g}(k) \bar{x}^H(k) \hat{R}_{xx}^{-1}(k-1) \quad (25-36)$$

Equation 25-36 is known as the Riccati equation for the RLS method. We can manipulate Eq. 25-35 to yield

$$\bar{g}(k) = \left[\alpha^{-1} \hat{R}_{xx}^{-1}(k-1) - \alpha^{-1} \bar{g}(k) \bar{x}^H(k) \hat{R}_{xx}^{-1}(k-1) \right] \bar{x}(k) \quad (25-37)$$

It is clear that the term inside the brackets of Eq. 25-37 is equal to Eq. 25-36. Thus

$$\bar{g}(k) = \hat{R}_{xx}^{-1}(k) \bar{x}(k) \quad (25-38)$$

Now we can derive a recursion relationship to update the weight vectors. The optimum Wiener solution is repeated next in terms of the iteration number k , and we can substitute Eq. 25-32 yielding

$$\begin{aligned} \bar{w}(k) &= \hat{R}_{xx}^{-1}(k) \hat{r}(k) \\ &= \alpha \hat{R}_{xx}^{-1}(k) \hat{r}(k-1) + \hat{R}_{xx}^{-1}(k) \bar{x}(k) d^*(k) \end{aligned} \quad (25-39)$$

We may now substitute Eq. 25-36 into the first covariance matrix inverse seen in Eq. 25-39:

$$\begin{aligned} \bar{w}(k) &= \hat{R}_{xx}^{-1}(k-1) \hat{r}(k-1) - \bar{g}(k) \bar{x}^H(k) \hat{R}_{xx}^{-1}(k-1) \hat{r}(k-1) + \hat{R}_{xx}^{-1}(k) \bar{x}(k) d^*(k) \\ &= \bar{w}(k-1) - \bar{g}(k) \bar{x}^H(k) \bar{w}(k-1) + \hat{R}_{xx}^{-1}(k) \bar{x}(k) d^*(k) \end{aligned} \quad (25-40)$$

Finally we may substitute Eq. 25-38 into Eq. 25-40 to yield

$$\begin{aligned} \bar{w}(k) &= \bar{w}(k-1) - \bar{g}(k) \bar{x}^H(k) \bar{w}(k-1) + \bar{g}(k) d^*(k) \\ &= \bar{w}(k-1) + \bar{g}(k) [d^*(k) - \bar{x}^H(k) \bar{w}(k-1)] \end{aligned} \quad (25-41)$$

Note that Eq. 25-41 is identical in form to Eq. 25-16.

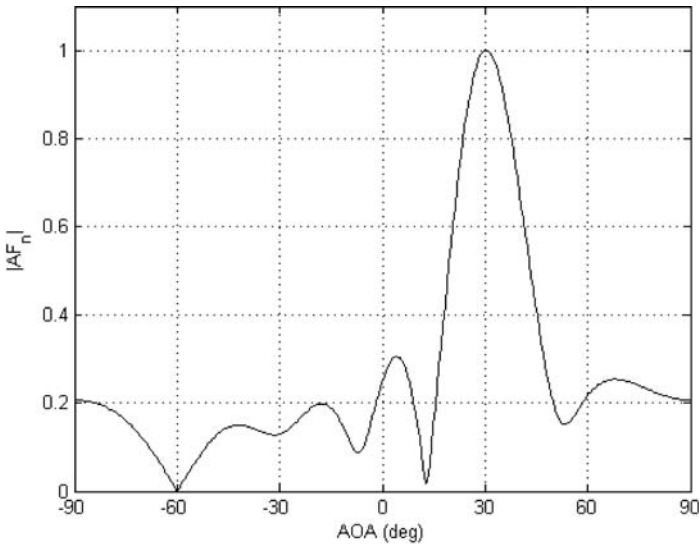


FIGURE 25-8 RLS array pattern

As an example of the RLS method, again let the array be an $M = 8$ element array with uniform spacing $\lambda/2$ with a desired signal arriving at the angle $\theta_0 = 30^\circ$ and interfering signal at $\theta_1 = -60^\circ$. As before, the noise variance is $\sigma_n^2 = .01$. Starting with initial array weights of zero and a forgetting factor $\alpha = .9$, Figure 25-8 shows the resulting RLS beam pattern.

The advantage of the RLS algorithm over SMI is that it is no longer necessary to invert a large covariance matrix. The recursive equations allow for easy updates of the inverse of the covariance matrix. The RLS algorithm also converges much more quickly than the LMS algorithm.

Conjugate Gradient Method

The problem with the steepest descent method is the sensitivity of the convergence rates to the eigenvalue spread of the covariance matrix. Greater spreads result in slower convergences. The convergence rate can be accelerated by use of the conjugate gradient method (CGM). The goal of CGM is to iteratively search for the optimum solution by choosing conjugate (perpendicular) paths for each new iteration. “Conjugacy” in this context is intended to mean “orthogonality.” The method of CGM produces orthogonal search directions resulting in the fastest convergence. Figure 25-9 depicts a top view of a two-dimensional performance surface where the conjugate steps show convergence toward the optimum solution. Note that the path taken at iteration $k + 1$ is perpendicular to the path taken at the previous iteration k .

The CGM method has its roots in the desire to solve a system of linear equations. One of the earliest references to CGM is found in a 1952 journal article by Hestenes and Stiefel.³⁶ Additionally, early work was performed by Fletcher and Powell³⁷ in 1963 and Fletcher and Reeves³⁸ in 1964. The CGM has also been called the *accelerated gradient* approach (AG) by Monzingo and Miller.²⁰ The CGM method was modified for use in predicting array weights by Choi et al.^{39,40} A survey of this method has been written by Godara,³³ and a

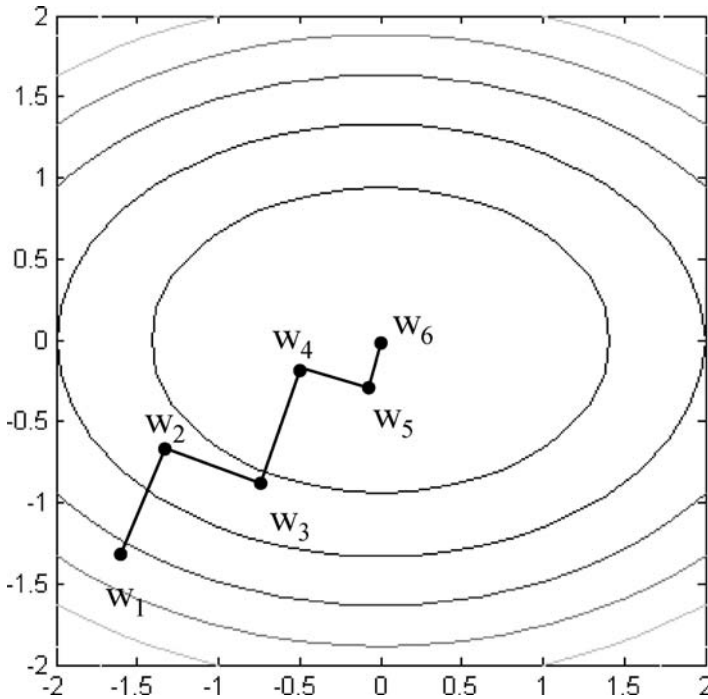


FIGURE 25-9 Convergence using conjugate directions

concise summary of the method can be found in Sadiku.⁴¹ The following summary is taken from the previous two references.

CGM is an iterative method whose goal is to minimize the quadratic cost function

$$C(\bar{w}) = \frac{1}{2} \bar{w}^H \bar{A} \bar{w} - \bar{d}^H \bar{w} \quad (25-42)$$

where

$$\bar{A} = \begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_M(1) \\ x_1(2) & x_2(2) & \cdots & x_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(K) & x_2(K) & \cdots & x_M(K) \end{bmatrix} \quad K \times M \text{ matrix of array snapshots}$$

K = number of snapshots

M = number of array elements

\bar{w} = unknown weight vector

$\bar{d} = [d(1) \ d(2) \ \dots \ d(K)]^T$ = desired signal vector of K snapshots

We may take the gradient of the cost function and set it to zero in order to find the minimum. It can be shown that

$$\nabla_{\bar{w}} C(\bar{w}) = \bar{A} \bar{w} - \bar{d} \quad (25-43)$$

We may employ the method of steepest descent in order to iterate toward the minimum of Eq. 25-43. We wish to slide to the bottom of the quadratic cost function by choosing an optimal route that requires the least number of iterations. Let us start with an initial guess for the weights $\bar{w}(1)$ and find the residual $\bar{r}(1)$. The first residual value after the first guess is given as

$$\bar{r}(1) = -\nabla_{\bar{w}} C(\bar{w}(1)) = \bar{d} - \bar{A}\bar{w}(1) \quad (25-44)$$

We can next choose a direction vector \bar{D} that gives us the new conjugate direction to iterate toward the optimum weight. Thus

$$\bar{D}(1) = \bar{A}^H \bar{r}(1) \quad (25-45)$$

The general weight update equation is given by

$$\bar{w}(n+1) = \bar{w}(n) - \mu(n) \bar{D}(n) \quad (25-46)$$

where the step size is determined by

$$\mu(n) = \frac{\bar{r}^H(n) \bar{A} \bar{A}^H \bar{r}(n)}{\bar{D}^H(n) \bar{A}^H \bar{A} \bar{D}(n)} \quad (25-47)$$

We may now update the residual and the direction vector. We can premultiply Eq. 25-46 by $-\bar{A}$ and add \bar{d} to derive the updates for the residuals:

$$\bar{r}(n+1) = \bar{r}(n) + \mu(n) \bar{A} \bar{D}(n) \quad (25-48)$$

The direction vector update is given by

$$\bar{D}(n+1) = \bar{A}^H \bar{r}(n+1) - \alpha(n) \bar{D}(n) \quad (25-49)$$

We can use a linear search to determine $\alpha(n)$, which minimizes $C(\bar{w}(n))$. Thus

$$\alpha(n) = \frac{\bar{r}^H(n+1) \bar{A} \bar{A}^H \bar{r}(n+1)}{\bar{r}^H(n) \bar{A} \bar{A}^H \bar{r}(n)} \quad (25-50)$$

The GCM procedure is to find the residual, the corresponding array weights, and to update until convergence is satisfied. It can be shown that the true solution can be found in no more than K iterations. This condition is known as *quadratic convergence*.

For the $M = 8$ element array with elements uniformly spaced $\lambda/2$, let us find the array weights under the following conditions: The arriving signal appears at 45° ; one interfering signal arrives at -30° and a second interfering signal arrives at 0° . The noise has a variance of $\sigma_n^2 = .001$, and we use a data block size $K = 20$. The plot of the norm of the residual is shown in Figure 25-10. It can be seen that the residual drops to very small levels after 14 iterations. The resulting antenna pattern is shown in Figure 25-11. Note that two nulls are placed at the two angles of arrival of the interference.

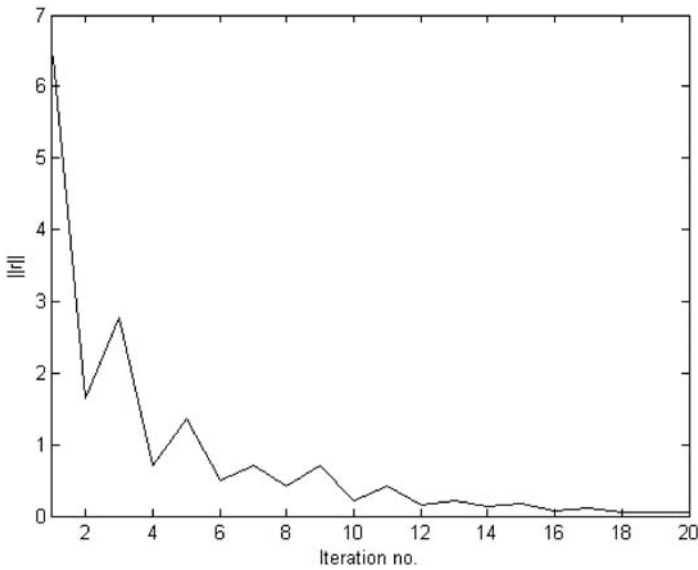


FIGURE 25-10 Norm of the residuals for each iteration

Constant Modulus

Many adaptive beamforming algorithms are based upon minimizing the error between a reference signal and the array output. The reference signal is typically a known sequence used to “train” the adaptive array or a desired signal based upon an a priori knowledge of

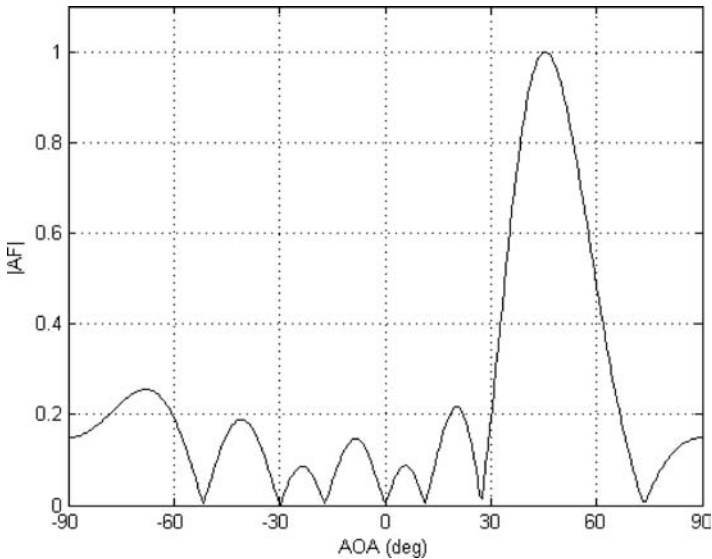


FIGURE 25-11 Array pattern using CGM

the arriving signals. In the case where a reference signal is not available, you must resort to an assortment of optimization techniques that are “blind” to the exact content of the incoming signals.

Many wireless communication and radar signals are frequency or phase modulated signals. Some examples of phase and frequency modulated signals are FM, PSK, FSK, QAM, and polyphase. Therefore the amplitude of the signal should ideally be a constant. Thus the signal is said to have a constant magnitude or *modulus*. However, in fading channels where multipath exists, the received signal is the composite of many delayed and attenuated replicas. Thus the propagation channel may introduce an amplitude variation on the signal magnitude. Frequency selective channels by definition destroy the constant modulus property of the signal. If we know that the arriving signals of interest should have a constant modulus, we can devise algorithms that restore or “equalize” the amplitude of the original signal.

Dominique Godard²³ was the first to capitalize on the constant modulus (CM) property in order to create a family of blind equalization algorithms used in two-dimensional data communication systems. Specifically, Godard’s algorithm is suited for phase modulating waveforms, using a cost function called a *dispersion* function of order p to determine optimum array element weights. The Godard cost function is given by

$$C(k) = E \left[\left(|y(k)|^p - R_p \right)^q \right] \quad (25-51)$$

where

$p, q = \text{positive integers}$

Godard showed that the gradient of the cost function is zero when R_p is defined by

$$R_p = \frac{E[|s(k)|^{2p}]}{E[|s(k)|^p]} \quad (25-52)$$

where

$s(k) = \text{zero-memory estimate of the array output } y(k)$

The resulting error signal is given by

$$e(k) = y(k) |y(k)|^{p-2} \left(R_p - |y(k)|^p \right) \quad (25-53)$$

This error signal can replace the traditional error signal in the LMS algorithm to yield

$$\bar{w}(k+1) = \bar{w}(k) + \mu e^*(k) \bar{x}(k) \quad (25-54)$$

The $p = 1, q = 2$ case reduces the cost function to the form

$$C(k) = E \left[\left(|y(k)| - R_1 \right)^2 \right] \quad (25-55)$$

where

$$R_1 = \frac{E[|s(k)|^2]}{E[|s(k)|]} \quad (25-56)$$

If we scale the array output estimate $s(k)$ to unity, we can write the error signal in Eq. 25-53 as

$$e(k) = \left(y(k) - \frac{y(k)}{|y(k)|} \right) \quad (25-57)$$

Thus the weight vector in the $p = 1$ case becomes

$$\bar{w}(k+1) = \bar{w}(k) + \mu \left(1 - \frac{1}{|y(k)|} \right) y^*(k) \bar{x}(k) \quad (25-58)$$

The $p = q = 2$ case reduces the cost function to the form

$$C(k) = E \left[\left(|y(k)|^2 - R_2 \right)^2 \right] \quad (25-59)$$

where

$$R_2 = \frac{E \left[|s(k)|^4 \right]}{E \left[|s(k)|^2 \right]} \quad (25-60)$$

If we scale the array output estimate $s(k)$ to unity, we can write the error signal in Eq. 25-53 as

$$e(k) = y(k) \left(1 - |y(k)|^2 \right) \quad (25-61)$$

Thus the weight vector, in the $p = 2$ case, becomes

$$\bar{w}(k+1) = \bar{w}(k) + \mu \left(1 - |y(k)|^2 \right) y^*(k) \bar{x}(k) \quad (25-62)$$

The cases where $p = 1$ or 2 are referred to as constant modulus algorithms (CMAs). The $p = 1$ case has been proven to converge much more rapidly than the $p = 2$ case.⁴² A similar algorithm was developed by Treichler and Agee⁴³ and is identical to the Godard case for $p = 2$.

Least Squares Constant Modulus

One severe disadvantage of the Godard CMA algorithm is the slow convergence time. The slow convergence limits the usefulness of the algorithm in dynamic environments where the signal must be captured quickly. This also limits the usefulness of CMA when propagation channel conditions are rapidly changing. The previous CMA method is based upon the method of steepest descent by taking the gradient of the cost function. A faster algorithm was developed by Agee²⁴ using the method of nonlinear least-squares. The least-squares method is also known as the Gauss method based upon the work of Gauss in 1795.⁴⁴ This method is referred to as the LS-CMA algorithm⁴⁵ and is also known as an autoregressive estimator based on a least squares minimization.⁴⁶

The following derivation is taken directly from Agee²⁴ and Rong.⁴⁵ In the method of least squares, one defines a cost function that is the weighted sum of error squares or the total error energy. The energies are the energies of a finite sample set K . The cost function is defined by

$$C(\bar{w}) = \sum_{k=1}^K |\phi_k(\bar{w})|^2 = \|\bar{\Phi}(\bar{w})\|_2^2 \quad (25-63)$$

where

$\phi_k(\bar{w})$ = error at k th data sample

$$\bar{\Phi}(\bar{w}) = [\phi_1(\bar{w}) \quad \phi_1(\bar{w}) \quad \cdots \quad \phi_K(\bar{w})]^T$$

K = number of data samples in one block

Eq. 25-63 has a partial Taylor-series expansion with a sum-of-squares form given as

$$C(\bar{w} + \bar{\Delta}) \approx \|\bar{\Phi}(\bar{w}) + \bar{J}^H(\bar{w})\bar{\Delta}\|_2^2 \quad (25-64)$$

where the complex Jacobian of $\bar{\Phi}(\bar{w})$ is defined as

$$\bar{J}(\bar{w}) = [\nabla\phi_1(\bar{w}) \quad \nabla\phi_2(\bar{w}) \quad \cdots \quad \nabla\phi_K(\bar{w})] \quad (25-65)$$

and

$\bar{\Delta}$ = offset, which updates weights

We wish to find the offset $\bar{\Delta}$ that minimizes the sum-of-squared errors. Taking the gradient of Eq. 25-64 and setting it equal to zero, we find the optimum offset vector to be

$$\bar{\Delta} = -[\bar{J}(\bar{w})\bar{J}^H(\bar{w})]^{-1}\bar{J}(\bar{w})\bar{\Phi}(\bar{w}) \quad (25-66)$$

The new updated weight vector is then given by

$$\bar{w}(n+1) = \bar{w}(n) - [\bar{J}(\bar{w}(n))\bar{J}^H(\bar{w}(n))]^{-1}\bar{J}(\bar{w}(n))\bar{\Phi}(\bar{w}(n)) \quad (25-67)$$

The new weight vector is the previous weight vector adjusted by the offset $\bar{\Delta}$. The number n is the iteration number not to be confused with the time sample k .

Let us now apply the least squares method to the constant modulus algorithm using the “1-2” cost function:⁸

$$C(\bar{w}) = \sum_{k=1}^K |\phi_k(\bar{w})|^2 = \sum_{k=1}^K \|y(k) - 1\|^2 \quad (25-68)$$

where

$$y(k) = \bar{w}^H \bar{x}(k) = \text{array output at time } k$$

We may write ϕ_k as a vector such that

$$\bar{\phi}(\bar{w}) = \begin{bmatrix} |y(1)| - 1 \\ |y(2)| - 1 \\ \vdots \\ |y(K)| - 1 \end{bmatrix} \quad (25-69)$$

We may now define the Jacobian of the error vector $\bar{\phi}(\bar{w})$ as

$$\begin{aligned} \bar{J}(\bar{w}) &= [\nabla(\phi_1(\bar{w})) \quad \nabla(\phi_2(\bar{w})) \quad \cdots \quad \nabla(\phi_K(\bar{w}))] \\ &= \begin{bmatrix} \bar{x}(1) \frac{y^*(1)}{|y(1)|} & \bar{x}(2) \frac{y^*(2)}{|y(2)|} & \cdots & \bar{x}(K) \frac{y^*(K)}{|y(K)|} \end{bmatrix} \\ &= \bar{X} \bar{Y}_{CM} \end{aligned} \quad (25-70)$$

where

$$\bar{X} = [\bar{x}(1) \quad \bar{x}(2) \quad \cdots \quad \bar{x}(K)] \quad (25-71)$$

and

$$\bar{Y}_{CM} = \begin{bmatrix} \frac{y^*(1)}{|y(1)|} & 0 & \cdots & 0 \\ 0 & \frac{y^*(2)}{|y(2)|} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{y^*(K)}{|y(K)|} \end{bmatrix} \quad (25-72)$$

Multiplying the Jacobian times its Hermitian transpose we get

$$\bar{J}(\bar{w}) \bar{J}^H(\bar{w}) = \bar{X} \bar{Y}_{CM} \bar{Y}_{CM}^H \bar{X}^H = \bar{X} \bar{X}^H \quad (25-73)$$

The product of the Jacobian times the energy matrix is given by

$$\bar{J}(\bar{w}) \bar{\Phi}(\bar{w}) = \bar{X} \bar{Y}_{CM} \begin{bmatrix} |y(1)| - 1 \\ |y(2)| - 1 \\ \vdots \\ |y(K)| - 1 \end{bmatrix} = \bar{X} \begin{bmatrix} y^*(1) - \frac{y^*(1)}{|y(1)|} \\ y^*(2) - \frac{y^*(2)}{|y(2)|} \\ \vdots \\ y^*(K) - \frac{y^*(K)}{|y(K)|} \end{bmatrix} = \bar{X}(\bar{y} - \bar{r})^* \quad (25-74)$$

where

$$\bar{y} = [y(1) \quad y(2) \quad \cdots \quad y(K)]^T \quad (25-75)$$

and

$$\bar{r} = \begin{bmatrix} \frac{y(1)}{|y(1)|} & \frac{y(2)}{|y(2)|} & \cdots & \frac{y(K)}{|y(K)|} \end{bmatrix}^T = L(\bar{y}) \quad (25-76)$$

In this form, note that $L(\bar{y})$ is a hard-limiter acting on \bar{y} . Substituting Eqs. 25-70 and 25-72 into Eq. 25-67 we get

$$\begin{aligned} \bar{w}(n+1) &= \bar{w}(n) - [\bar{X}\bar{X}^H]^{-1} \bar{X}(\bar{y}(n) - \bar{r}(n))^* \\ &= \bar{w}(n) - [\bar{X}\bar{X}^H]^{-1} \bar{X}\bar{X}^H \bar{w}(n) + [\bar{X}\bar{X}^H]^{-1} \bar{X}\bar{r}^*(n) \\ &= [\bar{X}\bar{X}^H]^{-1} \bar{X}\bar{r}^*(n) \end{aligned} \quad (25-77)$$

where

$$\bar{r}^*(n) = \begin{bmatrix} \frac{\bar{w}^H(n)\bar{x}(1)}{|\bar{w}^H(n)\bar{x}(1)|} & \frac{\bar{w}^H(n)\bar{x}(2)}{|\bar{w}^H(n)\bar{x}(2)|} & \cdots & \frac{\bar{w}^H(n)\bar{x}(K)}{|\bar{w}^H(n)\bar{x}(K)|} \end{bmatrix}^H \quad (25-78)$$

While only one block of data is used to implement the LS-CMA, the algorithm iterates through n values until convergence. Initial weights $\bar{w}(1)$ are chosen, the complex-limited output data vector $\bar{r}^*(1)$ is calculated, the next weight $\bar{w}(2)$ is calculated, and the iteration continues until satisfactory convergence is achieved. This is called the *static* LS-CMA algorithm because only one static block of length K is used for the iteration process. The LS-CMA algorithm bears a striking resemblance to the sample matrix inversion (SMI) algorithm in Eq. 25-26.

The chief advantage of the static LS-CMA is that it can converge up to two orders of magnitude faster than the conventional CMA algorithm. The static LS-CMA algorithm computes the weights simply based upon a fixed block of sampled data. To maintain up-to-date adaptation in a dynamic signal environment, it is better to update the data blocks per each iteration. Thus a dynamic LS-CMA algorithm is more appropriate. The dynamic LS-CMA is a modification of the previous static version. Let us define a dynamic block of data as the array output before applying weights. For the n th iteration, the n th block of length K is given as

$$\bar{X}(n) = [\bar{x}(1+nK) \quad \bar{x}(2+nK) \quad \cdots \quad \bar{x}(K+nK)] \quad (25-79)$$

The weighted array output for the n th iteration is now defined as

$$\bar{y}(n) = [y(1+nK) \quad y(2+nK) \quad \cdots \quad y(K+nK)]^T = [\bar{w}^H(n)\bar{X}(n)]^T \quad (25-80)$$

The complex limited output data vector is given as

$$\bar{r}(n) = \begin{bmatrix} \frac{y(1+nK)}{|y(1+nK)|} & \frac{y(2+nK)}{|y(2+nK)|} & \dots & \frac{y(K+nK)}{|y(K+nK)|} \end{bmatrix}^T \quad (25-81)$$

Replacing Eq. 25-77 with the dynamic version, we have

$$\bar{w}(n+1) = [\bar{X}(n)\bar{X}^H(n)]^{-1} \bar{X}(n)\bar{r}^*(n) \quad (25-82)$$

where

$$\bar{r}^*(n) = \begin{bmatrix} \frac{\bar{w}^H(n)\bar{x}(1+nK)}{|\bar{w}^H(n)\bar{x}(1+nK)|} & \frac{\bar{w}^H(n)\bar{x}(2+nK)}{|\bar{w}^H(n)\bar{x}(2+nK)|} & \dots & \frac{\bar{w}^H(n)\bar{x}(K+nK)}{|\bar{w}^H(n)\bar{x}(K+nK)|} \end{bmatrix}^H \quad (25-83)$$

We can further simplify Eq. 25-83 by defining the array covariance matrix and the covariance vector as

$$\hat{R}_{xx}(n) = \frac{\bar{X}(n)\bar{X}^H(n)}{K} \quad (25-84)$$

and

$$\hat{p}_{xr}(n) = \frac{\bar{X}(n)\bar{r}^*(n)}{K} \quad (25-85)$$

The dynamic LS-CMA is now defined as

$$\bar{w}(n+1) = \hat{R}_{xx}^{-1}(n)\hat{p}_{xr}(n) \quad (25-86)$$

As an example of the LS-CMA algorithm, let us define a constant modulus signal as a 32-chip binary sequence. Allow this signal to arrive at the receiver via a direct path and two additional indirect paths. Assume that the indirect multipaths are delayed a half-chip width and a full-chip width, respectively. Further, allow the first multipath term to be attenuated by 30 percent and the second path to be attenuated by 10 percent. The signal and multipath geometry is defined such that the direct path arrives at 45° to the array, the second path arrives at -30° , while the third path arrives at 0° . The SNR is 20 dB. Let us choose a block length of 24 data samples covering six chips. Allow the block to update for every iteration n . As in previous examples, we use an $M = 8$ element array with elements uniformly spaced at $\lambda/2$. Figure 25-12 shows the composite arriving signals before and after the application of the dynamic LS-CMA. Figure 25-13 shows the resulting array pattern.

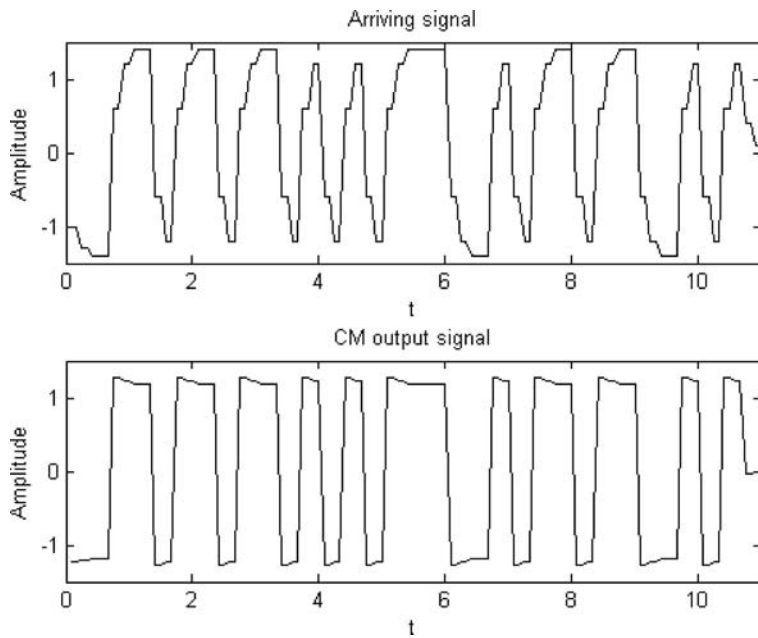


FIGURE 25-12 Composite arriving signal and the LS-CMA output signal

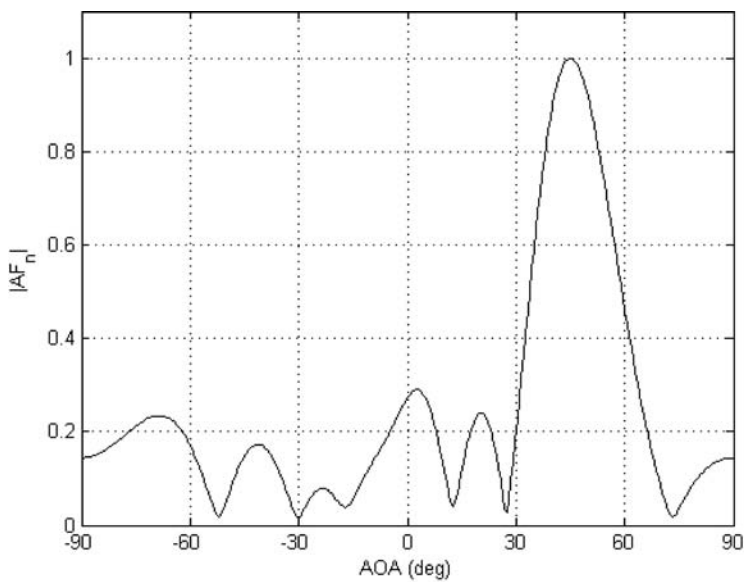


FIGURE 25-13 Dynamic LS-CMA array pattern

REFERENCES

1. P. Barton, "Digital Beamforming for Radar," *IEE Proc.*, Pt. F, vol. 127 (August 1980): 266–277.
2. E. Brookner, "Trends in Array Radars for the 1980's and Beyond," *IEEE Ant. and Prop. Society Newsletter* (April 1984): 27–30.
3. H. Steyskal, "Digital Beamforming Antennas—An Introduction," *Microwave Journal* (January 1987).
4. M. Skolnik, "System Aspects of Digital Beam Forming Ubiquitous Radar," Naval Research Lab, Report No. NRL/MR/5007—02-8625 (June 28, 2002).
5. M. Skolnik, *Introduction to Radar Systems*, 3rd Ed. (New York: McGraw-Hill, 2001).
6. T. Curtis, "Digital Beamforming for Sonar Systems," *IEE Proc.*, Pt. F, vol. 127 (August 1980): 257–265.
7. J. Litva and T. K.-Y. Lo, *Digital Beamforming in Wireless Communications* (Boston: Artech House, 1996).
8. J. Liberti and T. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications* (Upper Saddle River: Prentice-Hall, 1999).
9. F. Gross, *Smart Antennas for Wireless Communications with MATLAB* (New York: McGraw-Hill, 2005).
10. L. Van Atta, "Electromagnetic Reflection," (October 6, 1959): U.S. Patent 2908002.
11. D. Margerum, "Self-Phased Arrays," *Microwave Scanning Antennas*, Chap. 5, vol. III: *Array Systems*, R. C. Hansen (ed.) (New York: Academic Press, 1966).
12. R. York and T. Itoh, "Injection and Phase-Locking Techniques for Beam Control," *IEEE Trans. MTT*, vol. 46, no. 11 (November 1998): 1920–1929.
13. P. Howells, "Intermediate Frequency Sidelobe Canceller," (August 24, 1965): U. S. Patent 3202990.
14. P. Howells, "Explorations in Fixed and Adaptive Resolution at GE and SURC," *IEEE Trans. Antennas Propagat.*, Special Issue on Adaptive Antennas, vol. AP-24, no. 5 (September 1976): 575–584.
15. S. Applebaum, "Adaptive Arrays," Syracuse University Research Corporation, Rep. SPL TR66-1 (August 1966).
16. S. Applebaum, "Adaptive Arrays," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 5 (September 1976): 585–598.
17. B. Widrow and M. Hoff, "Adaptive Switch Circuits," *IRE Wescom, Convention Record*, Part 4 (1960): 96–104.
18. B. Widrow, P. Mantey, L. Griffiths, and B. Goode, "Adaptive Antenna Systems," *Proc. IEEE*, vol. 55 (December 1967).
19. W. Gabriel, "Adaptive Processing Antenna Systems," *IEEE Antennas and Propagat. Newsletter* (October 1983): 5–11.
20. R. Monzingo and T. Miller, *Introduction to Adaptive Arrays* (New York: Wiley-Interscience, John Wiley & Sons, 1980).
21. I. Reed, J. Mallett, and L. Brennen, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10 (November 1974): 853–863.
22. J. D. Powell, "An Iterative Method for Finding Stationary Values of a Function of Several Variables," *Comput. J.* vol. 5, no. 2 (July 1962): 147–151.
23. D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Trans. Comm.*, vol. COM-28, no. 11 (November 1980): 1867–1875.
24. B. Agee, "The Least-Squares CMA: A New Technique for Rapid Correction of Constant Modulus Signals," *IEEE International Conference on ICASSP '86*, vol. 11 (April 1986): 953–956.
25. J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," *IEEE Proc.*, vol. 57, no. 8 (August 1969): 1408–1418.
26. J. P. Burg, "The Relationship Between Maximum Entropy Spectra and Maximum Likelihood Spectra," *Geophysics*, vol. 37 (April 1972): 375–376.

27. J. P. Burg, "Maximum Entropy Spectrum Analysis," Ph.D. dissertation, Dept. of Geophysics, Stanford University, Stanford, CA, 1975.
28. V. F. Pisarenko, "The Retrieval of Harmonics from a Covariance Function," *Geophysical Journal of the Royal Astronomical Society*, vol. 33: (1973): 347–366.
29. S. S. Reddi, "Multiple Source Location—A Digital Approach," *IEEE Trans. AES*, vol. 15, no.1 (January 1979).
30. R. Kumaresan and D. Tufts, "Estimating the Angles of Arrival of Multiple Plane Waves," *IEEE Trans. AES*, vol. AES-19 (1983): 134–139.
31. R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, no. 2 (March 1986): 276–280.
32. R. Roy and T. Kailath, "ESPRIT—Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. on ASSP*, vol. 37, no. 7 (July 1989): 984–995.
33. L. Godara, *Smart Antennas* (Boca Raton: CRC Press, 2004).
34. S. Haykin, *Adaptive Filter Theory*, 4th Ed. (Upper Saddle River: Prentice Hall, 2002).
35. G. H. Golub and C. H. Van Loan, *Matrix Computations*, 3rd Ed. (Baltimore: The Johns Hopkins University Press, 1996).
36. M. Hestenes and E. Stiefel, "Method of Conjugate Gradients for Solving Linear Systems," *Journal of Research of the National Bureau of Standards*, vol. 49 (1952): 409–436.
37. R. Fletcher and M. Powell, "A Rapidly Convergent Descent Method for Minimization," *Computer Journal*, vol. 6 (1963): 163–168.
38. R. Fletcher and C. Reeves, "Function Minimization by Conjugate Gradients," *Computer Journal*, vol. 7 (1964): 149–154.
39. S. Choi, *Application of the Conjugate Gradient Method for Optimum Array Processing*, Book Series on PIER (Progress in Electromagnetics Research), vol. 5 (Amsterdam: Elsevier, 1991).
40. S. Choi and T. Sarkar, "Adaptive Antenna Array Utilizing the Conjugate Gradient Method for Multipath Mobile Communication," *Signal Processing*, 29 (1992): 319–333.
41. M. Sadiku, *Numerical Techniques in Electromagnetics*, 2nd Ed., (Boca Raton: CRC Press, 2001).
42. M. Larimore and J. Treichler, "Convergence Behavior of the Constant Modulus Algorithm, Acoustics," *IEEE International Conference on ICASSP '83*, vol. 8 (April 1983): 13–16.
43. J. Treichler and B. Agee, "A New Approach to Multipath Correction of Constant Modulus Signals," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no.2 (April 1983): 459–472.
44. H. Sorenson, "Least-Squares Estimation: from Gauss to Kalman," *IEEE Spectrum*, vol. 7 (July 1970): 63–68.
45. Z. Rong, "Simulation of Adaptive Array Algorithms for CDMA Systems," Master's Thesis MPRG-TR-96-31, Mobile & Portable Radio Research Group, Virginia Tech, Blacksburg, VA, September 1996.
46. P. Stoica and R. Moses, *Introduction to Spectral Analysis* (Upper Saddle River: Prentice-Hall, 1997).

Chapter 26

Methods of
Polarization
Synthesis

Nathan Cummings
Sciperio, Inc.

CONTENTS

26.1	GENERAL DISCUSSION	22-2
26.2	POLARIZATION-SPECIFIC APPLICATIONS.	22-9
26.3	ANTENNA POLARIZATION AND POLARIZATION SYNTHESIS	22-12
26.4	DUAL-MODE HORN RADIATORS	22-15
26.5	MICROSTRIP PATCH RADIATORS.	22-21
26.6	TRANSMISSION- AND REFLECTION-TYPE POLARIZERS	22-22
26.7	POLARIZATION MEASUREMENTS	22-24

26.1 GENERAL DISCUSSION

The plane of polarization, or simply the polarization, of a radio wave is defined by the direction in which the electric field vector is aligned during the passage of at least one full cycle. In the general case, both the magnitude and the orientation of the electric field vector will vary during each cycle, and the electric field vector will map out an ellipse in the plane normal to the direction of propagation at the point of observation. In this general case (see Figure 26-1a), the polarization of the wave is said to be elliptical. The minor-to-major-axis ratio of the ellipse is called the *ellipticity* and will be expressed in this chapter in decibels. (Although the axis ratio is less than unity, when expressing ellipticity in decibels, the minus sign is frequently omitted for convenience. The term *axial ratio* is also in common use. It is the reciprocal of ellipticity.) The direction in which the major axis lies is called the *polarization orientation* and in this chapter will be measured from the vertical (see Figure 26-2).

The two special cases of ellipticity of particular interest are (1) an ellipticity of ∞ dB (minor-to-major-axis ratio zero), which is *linear polarization*, and (2) an ellipticity of 0 dB (minor-to-major-axis ratio unity), which is *circular polarization*. A linearly polarized wave is therefore defined as a transverse electromagnetic wave whose electric field vector (at a point in a homogeneous isotropic medium) at all times lies along a fixed line. A circularly polarized wave is similarly defined as a transverse electromagnetic wave for which the electric and/or magnetic field vector at a point describes a circle. In attempting to produce a linearly polarized wave, elliptical polarization is thought of as imperfect linear polarization, while in attempting to produce a circularly polarized wave, elliptical polarization is thought of as imperfect circular polarization.

Confusion occasionally results in the use of mental pictures similar to Figure 26-1 and Figure 26-2 when you overlook the fact that although the electric vector makes one complete revolution (see Figure 26-2) per cycle, it does not rotate at a uniform rate except in the special case of a circularly polarized wave. In this special case, rotation occurs at the rate of ω rad/s.

Figure 26-1c shows a circularly polarized wave having a right-hand sense. It is also possible, of course, to have left-hand circularly polarized waves. The definition of right-hand circular polarization as standardized¹ by the Institute of Electrical and Electronics Engineers and as used in this chapter is as follows: for an observer looking in the direction of propagation, the rotation of the electric field vector in a stationary transverse plane is clockwise for right-hand polarization. Similarly, the rotation is counterclockwise for left-hand polarization.

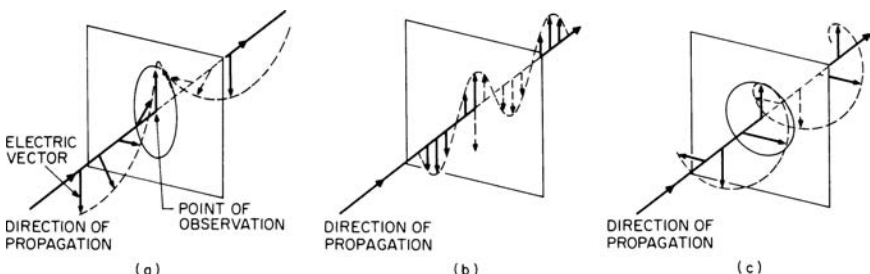


FIGURE 26-1 Diagrammatic illustration of waves of various polarization: (a) Elliptical polarization; (b) Linear polarization; (c) Circular polarization (right-hand)

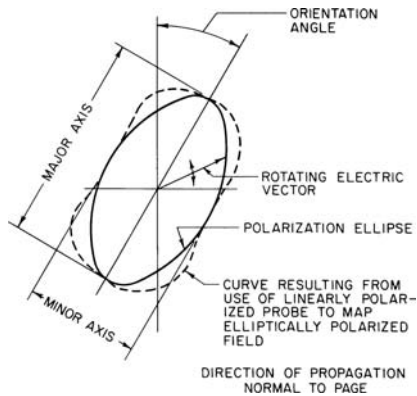


FIGURE 26-2 Polarization ellipse

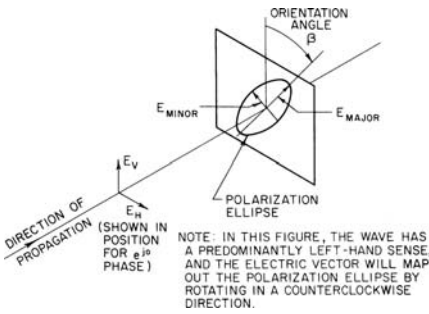
One simple way of determining experimentally the sense of rotation of a circularly polarized wave is to make use of two helical beam antennas of opposite sense. A right-hand helical antenna transmits or receives right-hand polarization, while a left-hand helical antenna transmits or receives left-hand polarization. If a circularly polarized wave is received first on a right-hand helical antenna and then on a left-hand helical antenna, the antenna that receives the greater amount of signal will have a sense that corresponds to the sense of the received wave. In the case of an elliptically polarized wave, the sense will be taken to be the same as that of the predominant circular component.

It is important to note that not all workers in the field use the same definition of polarization sense, and confusion may arise as a result of the ambiguity in its definition. The work of different authors should be compared with this precaution in mind.

Wave Polarization State Synthesis

Any wave of arbitrary polarization can be synthesized from two waves orthogonally polarized to each other. For example, a circularly polarized wave will be produced by the coexistence of a vertically and a horizontally polarized wave, each having the same amplitude and with a 90° phase difference between them. If they have other than the same amplitude and/or other than a 90° relationship, the resulting wave will be elliptically polarized. If, for example, the amplitude of the vertically polarized wave is zero, the resulting wave is linearly polarized and has a horizontal orientation. Further, if the two waves have equal amplitude but 0° phase difference, the resulting wave is linearly polarized with 45° orientation. There are two possible combinations of vertically and horizontally polarized waves that can produce a wave of some specific ellipticity. One of the combinations will produce a predominantly left-hand wave, and the other combination will produce a predominantly right-hand wave. Figures 26-3 and 26-4 show the two cases. Separate illustrations are provided to emphasize the need for care in specifying the coordinate system in this work and the need for consistency in adhering to the selected coordinate system while performing and interpreting any theoretical or experimental work.

It is often of interest to know only the ratio of the magnitudes of the vertically and horizontally polarized waves forming an elliptically polarized wave and the phase angle



$$E_V = \frac{1}{2} [E_{\text{MAJOR}} (1 + e^{j2\beta}) + E_{\text{MINOR}} (1 - e^{j2\beta})]$$

$$E_H = \frac{e^{j\pi/2}}{2} [E_{\text{MAJOR}} (1 - e^{j2\beta}) + E_{\text{MINOR}} (1 + e^{j2\beta})]$$

FIGURE 26-3 Linear-component synthesis of an elliptically polarized wave having a predominantly left-hand sense

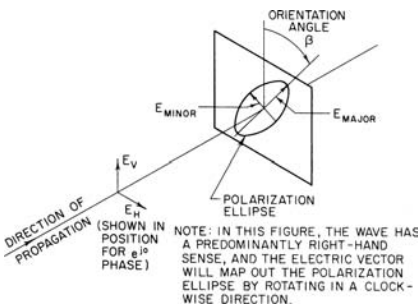
between them. Such information can be obtained graphically with the aid of Figure 26-5. In this chart, the phase angle shown is the relative phase of the vertical element when the relative phase of the horizontal element is zero.

Example Given an ellipticity of 3.0 dB and an orientation of $\beta = 37^\circ$, find the ratio of the magnitudes of the vertically and horizontally polarized waves and the phase angle between them. Draw a line from the center of the chart to the periphery at the 37° point. Lay off a distance from the chart center to the $V/H = 3.0$ -dB line measured on the horizontal diameter. Read $V/H = 0.8$ dB and a phase angle of 71° for predominantly right-hand sense or 289° for predominantly left-hand sense.

It is also possible to synthesize any elliptically polarized wave from two circularly polarized waves having opposite senses. For example, a linearly polarized wave will be produced by the coexistence of a right-hand and left-hand circularly polarized wave of the same amplitude. The orientation of the resulting linearly polarized wave will be determined by the phase difference between the two circularly polarized waves. Figure 26-6 provides the formulas for calculating the values.

Characteristics of Wave Polarization

Any discussion of wave polarization typically begins with a brief description including a few of the unique physical characteristics associated with electromagnetic wave propagation and wave interaction with the surrounding environment. It is important to note that while a complete description of wave polarization can require significant mathematical development, the resulting observable wave behavior can be expressed by several accessible characteristics.



$$E_V = \frac{1}{2} [E_{\text{MAJOR}} (1 + e^{j2\beta}) - E_{\text{MINOR}} (1 - e^{j2\beta})]$$

$$E_H = \frac{e^{j\pi/2}}{2} [E_{\text{MAJOR}} (1 - e^{j2\beta}) - E_{\text{MINOR}} (1 + e^{j2\beta})]$$

FIGURE 26-4 Linear-component synthesis of an elliptically polarized wave having a predominantly right-hand sense

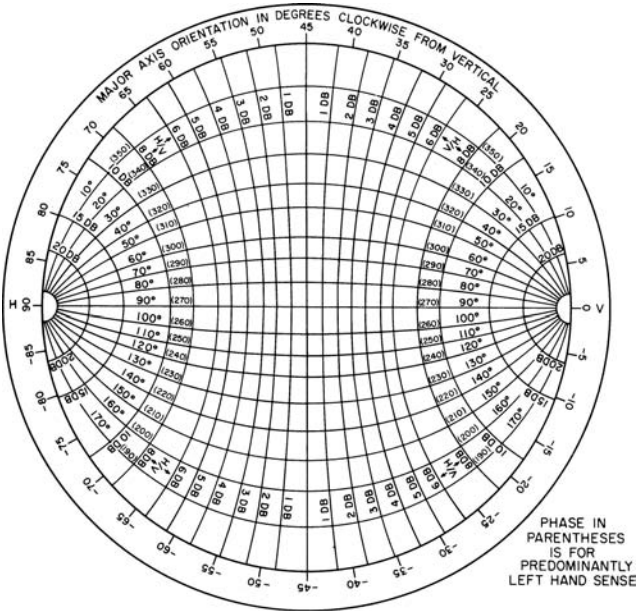


FIGURE 26-5 Polarization chart

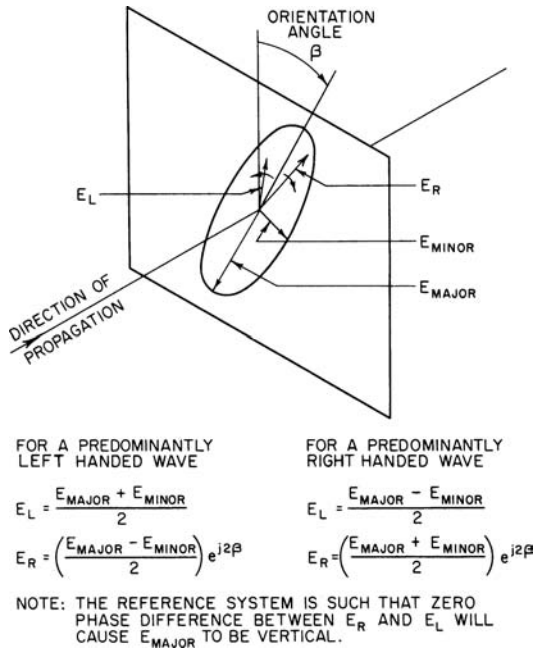


FIGURE 26-6 Circular-component synthesis of an elliptically polarized wave

TABLE 26-1 Two Sample Applications of Circularly Polarized Antennas and Some Typical Characteristics

	Response to Linear Polarization of Arbitrary Orientation	Precipitation-clutter Suppression in Radar Service (values based on search radar)
Ideal ellipticity*	0 dB	0 dB
Satisfactory ellipticity*	−3 dB	−0.5 dB
Unsatisfactory ellipticity*	−10 dB	−1.5 dB
Azimuth beamwidth	Omnidirectional	1–5°
Elevation beamwidth	20°	20°
Bandwidth	Greater than 50 percent	5–15 percent
Standing-wave ratio	5	1.15
Power rating	Milliwatt level	>100kW, <10MW

*Over all or almost all of the radiation pattern

Polarization Quality Although there is reasonably general agreement on the preceding definitions, the use of the words *circular polarization* does not always have quite the same meaning to different workers in the field. The difference lies in the permissible departure (for the application at hand) from precise polarization circularity before the circularly polarized antenna becomes unsatisfactory. The application of a circularly polarized antenna often falls into one of two categories. The first category is the use of a circularly polarized antenna to provide response to a linearly polarized wave of arbitrary orientation. The second category is the use of a circularly polarized antenna for suppression of precipitation clutter in radar service.² Table 26-1 shows typical characteristics for each of the applications.

Orthogonal circularly polarized antenna pairs are also in use for frequency-reuse applications with communications satellites.³ Channel isolation in excess of 20 dB is readily attainable. Stutzman and Overstreet⁴ describe a practical polarization-ellipticity field measurement technique. Further discussion is in Section 26.2.

Orthogonality For any arbitrarily polarized antenna, there can be another antenna polarized so that it will not respond to the wave emanating from the first antenna. The polarizations of the two antennas are said to be orthogonal. Using the convention of Figure 26-6, two polarization ellipses will be orthogonal if they satisfy the relation

$$-E_{R1} E_{R2} = E_{L1} E_{L2} \tag{26-1}$$

Figure 26-5 can be used to determine the orthogonal-polarization ellipse to a given ellipse by performing an inversion through the center of the chart and reading the phase-shift scale associated with the opposite predominant sense.

Reflection When a vertically polarized wave is reflected from a smooth surface, there is no change in its character.* When a horizontally polarized wave is reflected from a smooth

*Strictly speaking, this statement is correct only if the smooth surface is the interface between a normal propagation medium and one having an impedance of infinity. However, we are here interested only in the relative difference between the vertical and horizontal cases, so the detail will be overlooked.

surface, there is a 180° phase shift because of the coordinate system reversal in space when you look in the reversed direction of propagation. When a circularly polarized wave is reflected from a smooth surface, its horizontal component is altered by 180° ; hence the sense of its polarization is reversed. For an elliptically polarized wave, reflection is equivalent to altering the differential phase shift (e.g., the phase difference between the horizontally and vertically polarized linear components) by 180° . The new polarization ellipse may be determined with the aid of Figure 26-5 by inverting the original (before reflection) ellipticity across the horizontal diameter and reading the phase-shift scale corresponding to the opposite predominant sense.

Circularly polarized antennas are unique in being entirely unable to “see” their own images in any symmetrical reflecting surface, since the reflected wave has its sense reversed and is, therefore, orthogonal to the polarization of the antenna from which it originated.

Polarization Transmission Efficiency Within a communication link consisting of a transmitting antenna and a receiving antenna, it is useful to define a factor to quantify how well the receiving antenna extracts energy from the transmitting antenna with respect to the transmitted wave polarization. Polarization transmission efficiency or simply *polarization efficiency*, p , is the ratio of power accepted by a receiving antenna of a given polarization to the power accepted from the same transmitted wave passing an antenna whose polarization has been adjusted to match the incoming wave polarization. In order that a receiving antenna may extract the maximum amount of energy from the passing radio wave, it must have a polarization identical to that of the passing wave. For example, a vertically polarized receiving antenna should normally be used to receive a signal from a vertically polarized transmitting antenna assuming no polarization distortion in the transmission path. Similarly, a right-hand circularly polarized antenna should be used for the reception of waves from a right-hand circularly polarized transmitting antenna. In general, maximum transmission will result between two elliptically polarized antennas when:

1. Their axis ratios $E_{\text{minor}}/E_{\text{major}}$ are the same.
2. Their predominant senses are the same.
3. Their ellipse orientations are translated by a minus sign; that is, $\beta_{\text{trans}} = -\beta_{\text{rec}}$.

Condition 3 will result in a parallel-ellipse major axis in space and implies that only in three special cases will identical antennas at each end of the circuit with the same orientation yield maximum transmission. These three cases are

1. $\beta = 0^\circ$ or 180°
2. $\beta = \pm 90^\circ$
3. $E_{\text{minor}}/E_{\text{major}} = 1$, so that β has no significance

Identical antennas can always be made to yield maximum transmission, however, if they are rotated to cause the major axes of their waves to be parallel in space.

When two arbitrarily polarized antennas are used, the polarization efficiency from the receiving-antenna terminals will be

$$p = \frac{|(E_{R1}E_{R2} + E_{L1}E_{L2})|^2}{(|\mathbf{E}_{R1}|^2 + |\mathbf{E}_{R2}|^2)(|\mathbf{E}_{R2}|^2 + |\mathbf{E}_{L2}|^2)} \quad (26-2)$$

All four variables in Eq. 26-2 are vector quantities. Similar expressions in terms of linear components, or axial ratios, may be found in Sichak and Milazzo⁵ and Hatkin.⁶

TABLE 26-2 Polarization Transmission Efficiency with Various Polarizations at the Transmitting and Receiving Antennas

Antenna 1	Antenna 2	p
Vertical, e.g.: $E_R = 1$ $E_L = 1e^{j0} \quad \beta = 0$	Vertical, e.g.: $E_R = 1$ $E_L = 1e^{j0} \quad \beta = 0$	1
Vertical, e.g.: $E_R = 1$ $E_L = 1e^{j0} \quad \beta = 0$	Horizontal, e.g.: $E_R = 1$ $E_L = 1e^{j\pi} \quad \beta = 90^\circ$	0
Vertical, e.g.: $E_R = 1$ $E_L = 1e^{j0} \quad \beta = 0$	Circular (right), e.g.: $E_R = 1$ $E_L = 0$	$\frac{1}{2}$
Circular (right), e.g.: $E_R = 1$ $E_L = 0$	Circular (left), e.g.: $E_R = 0$ $E_L = 1$	0
Circular (left), e.g.: $E_R = 0$ $E_L = 1$	Circular (left), e.g.: $E_R = 0$ $E_L = 1$	1

The values of E_R and E_L may be obtained with the aid of Figure 26-6. Table 26-2 shows some typical combinations of interest.

Polarization efficiency is frequently expressed in dB and referred to as *polarization loss* or *polarization mismatch factor*. Polarization loss is given by

$$L_p = -10 \log(p) \tag{26-3}$$

Figure 26-7 shows a graphical means developed by Ludwig to display the maximum and minimum values of loss due to mismatched polarization between two antennas of arbitrary polarization. As an example of the use of the graph, if a predominantly right-hand polarized (RCP) transmitting antenna has an axial ratio of 8 dB and a predominantly left-hand polarized (LCP) receiving antenna has an axial ratio of 4 dB, draw two straight lines that intersect at the point P shown on the graph. This point is near the minimum-loss curve of 4 dB and the maximum-loss curve of 16 dB, and therefore the loss will vary (approximately) between 4 and 16 dB as one or the other antenna is rotated about its axis.* Further use of the graph to determine true gain with respect to a perfectly polarized antenna is described in Ludwig.⁷

Depolarization and Anisotropic Media

The polarization efficiency between two like polarized antennas in a communication link can be reduced through wave depolarization. *Depolarization* is the deviation of the wave polarization from its intended state. Normally, the polarization state of a wave traveling through isotropic media remains constant along the direction of propagation. Many common media types including air are considered to be nominally isotropic. However, a wave

* Portions of this paragraph and Figure 23-7 are reprinted with permission from the publishers of *Microwave Journal*.

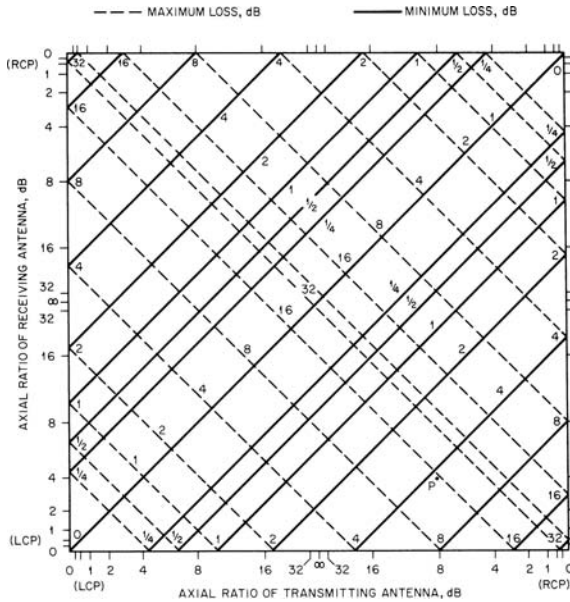


FIGURE 26-7 Polarization loss between two elliptically polarized antennas

propagating through a media that is *anisotropic* may have its polarization state altered as it passes through the media. Ionized gasses and crystalline structures are typical media responsible for wave depolarization.

One common occurrence of depolarization due to anisotropic media occurs in transionospheric satellite links.⁸ The earth's ionosphere contains partially ionized gas, and the presence of the earth's magnetic field causes the ionosphere to become anisotropic. The effect known as Faraday rotation occurs as a wave passes through the ionosphere and undergoes a rotation of its polarization orientation angle. This rotation results in the depolarization of linearly polarized signals. Faraday rotation is frequency- and path-length-dependent and is detailed further in Section 54.7. Circularly polarized signals are often used in satellite communication links to mitigate the impact of Faraday rotation caused by the ionosphere.

26.2 POLARIZATION-SPECIFIC APPLICATIONS

Frequency Reuse by Polarization Decoupling

The expanded number of communications satellites in use has made it necessary to use the same frequency to communicate with closely spaced areas on the earth. Commonly called *frequency reuse*, this method requires a decoupling mechanism between overlapping beams. Although careful shaping of the beams has been used in some cases, the use of orthogonal polarizations is more common. However, this puts stringent requirements on the amount of cross-polarization that can be present on any individual beam. Since the most common type of antenna used for satellite communications is a reflector antenna, the polarization properties of reflectors have been studied in great detail.⁹⁻¹³

Offset reflectors, which essentially eliminate aperture blockage, in particular have received a considerable amount of attention. Methods of computing the interbeam isolation and thus the frequency-reuse performance have been developed.¹⁴ For a symmetrical paraboloid, it has been shown that a physically circular feed with equal E - and H -plane amplitude and phase patterns will produce no cross-polarization;¹⁵ however, in an offset configuration, cross-polarization will be present because of the asymmetry. Chu and Turrin¹² have computed the maximum cross-polarization as a function of the offset angle θ_c and half of the reflector subtended angle θ_0 for linear polarization with a circular symmetric feed providing a 10-dB taper at the aperture edges. Their results, presented in Figures 26-8 and 26-9, show that the amount of cross-polarization decreases with increasing f/D and increases with increasing offset angle. In addition, they have shown that if the feed is perfectly circularly polarized, no cross-polarized energy will be found in the far-field pattern, although the position of the beam will be displaced from the axial direction because of a phase shift across the aperture. Figure 26-10 presents their results as a function of θ_0 and θ_c as a parameter.

The reduction of cross-polarized energy in offset configurations can be accomplished by careful feed design. The complex focal fields for an incoming plane wave can be calculated and a feed horn designed to match the focal plane by using multimode or hybrid

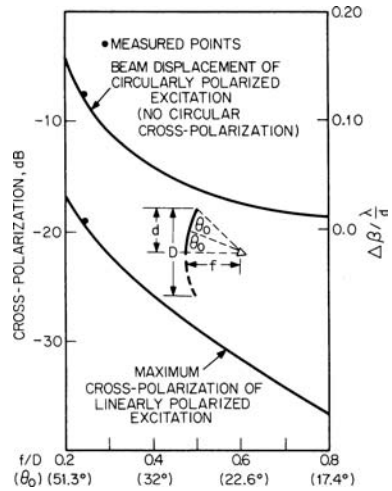


FIGURE 26-8 Cross-polarization and beam displacement versus f/D ratio ($\theta_0 = \theta_c$) (after T. S. Chu and R. H. Turrin¹² © IEEE 1973)

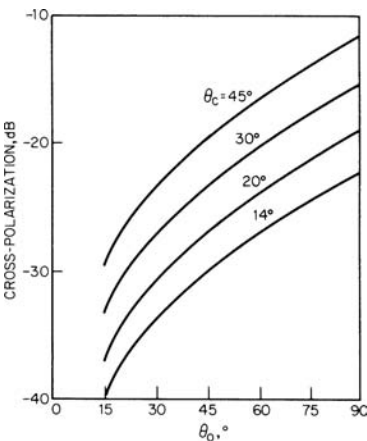


FIGURE 26-9 Maximum cross-polarization of linearly polarized excitation (after T. S. Chu and R. H. Turrin¹² © IEEE 1973)

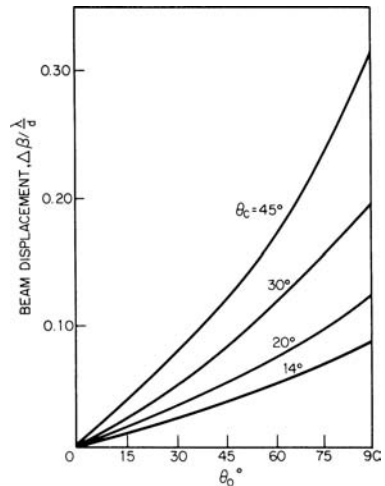


FIGURE 26-10 Beam displacement of circularly polarized excitation (no circular cross-polarization) (T. S. Chu and R. H. Turrin¹² © IEEE 1973)

modes (see Chapter 14). Alternatively, a polarizing grid can be used in front of the feed horn to reduce the cross-polarization.¹⁶ In general, the grid will not consist of parallel wires but will be aligned to match the polarization of the primary field incident on the reflector surface. Additional details on the subject of frequency reuse are found in Chapters 15, 44, 45, and 54.

Radar Precipitation-Clutter Suppression

Precipitation clutter is the name given to the radar echo from such targets as rain, snow, etc. It is not unusual for a radar to be rendered useless because precipitation clutter is sufficiently strong and extensive to hide the presence of a desired target (such as an aircraft). Because raindrops are substantially spherical (or at least much more so than an aircraft), they qualify as symmetrical reflectors. As mentioned in Section 26.1, a circularly polarized antenna is unable to see its own image in a symmetrical reflector. Therefore, if the radar antenna is circularly polarized, the echo from a symmetrical target such as a spherical raindrop will be circularly polarized with the wrong sense to be accepted by the antenna and will not be received. The echo from a composite target such as an aircraft will have scrambled polarization and will usually contain a polarization component to which the circularly polarized radar antenna can respond.

In other words, if the radar antenna is perfectly circularly polarized and the clutter is caused by spherical raindrops, the clutter cancellation will be perfect, enabling the presence of an otherwise-hidden composite target to be detected. When the radar antenna is elliptically polarized, the cancellation of an echo from symmetrical targets will not be complete; also, if the antenna is circularly polarized and the target is not symmetrical, cancellation will not be complete.

Cancellation ratio (CR) is defined as the ratio of radar power received from a symmetrical target when using a precipitation-clutter-suppression technique to the power received from the same target when not using the suppression technique. Cancellation ratio is not the same as cross-polarization ratio. Figure 26-11 shows cancellation ratio as a function of the polarization ellipticity of the antenna illuminating the symmetrical target.

Cancellation ratio, expressed in decibels, is related to the antenna polarization ellipticity by

$$CR_{dB} = 20 \log \left[\frac{1 - E^2}{1 + E^2} \right] \quad (26-4)$$

Note that the polarization ellipticity E is the ratio of minor to major axis (field quantities) and that the bracketed fraction is a voltage ratio.

Table 26-4 shows the extent of cancellation of various targets when the antenna is perfectly circularly polarized.

The use of circular polarization results in some reduction in the radar power received from an aircraft target also. The reduction can vary over wide limits, depending on aircraft type, aspect ratio, wavelength, etc., and by its nature it can only be described statistically. However, reduction values well under 10 dB are often observed. Additionally, dense precipitation usually does not completely surround a radar at all ranges, but tends to occur in patches. The net improvement in apparent

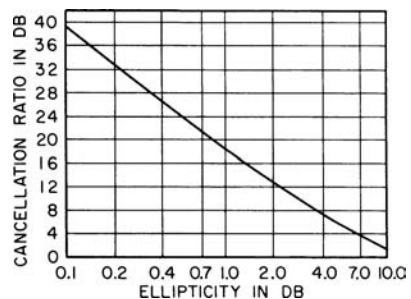


FIGURE 26-11 Cancellation ratio as a function of polarization ellipticity

TABLE 26-4 Cancellation of Various Targets for Circular Polarization

Target	Return Using Circular Polarization Relative to Return Using Linear Polarization
Sphere	Complete cancellation
Disk facing radar	Complete cancellation
Large sheet facing radar	Complete cancellation
Wire grating facing radar (wires parallel to linear polarization for that case)	6-dB cancellation
Double-bounce corner reflector	Zero cancellation
Triple-bounce corner reflector	Complete cancellation

signal-to-noise ratio (that is, the desired aircraft-target signal to precipitation-clutter noise) obtained by using circular polarization for precipitation-clutter suppression is worthwhile, with net improvement in the 10- to 20-dB area being readily achievable.

26.3 ANTENNA POLARIZATION AND POLARIZATION SYNTHESIS

Circular polarization can be achieved by a combination of electric and magnetic antennas provided that the fields produced by these antennas are equal in magnitude and in time-phase quadrature. A simple case of this combination is a horizontal loop and a vertical dipole.¹⁷ The time-phase-quadrature relationship is a fundamental relationship between the fields of a loop and a dipole when their currents are in phase. If the loop and dipole are oriented as in Figure 26-12, the fields in the plane of the loop are given by

$$E_L = jCJ_1(ka)e^{j(\omega t - kr)} \tag{26-5}$$

$$E_D = C_1e^{j(\omega t - kr)} \tag{26-6}$$

where

- C and C_1 = constants
- $k = 2\pi/\lambda$
- λ = free-space wavelength
- a = radius of loop
- r = distance from center of loop
- J_1 = Bessel function of the first order

provided that the currents in the loop and dipole are in phase. Thus, if

$$CJ_1(ka) = C_1 \tag{26-7}$$

the resulting field of the combination will be circularly polarized. Equation 26-7 will be true if the loop diameter is less than about 0.6 wavelength and the dipole length is less than a half wavelength. In this particular combination it should be noted that the resulting radiation

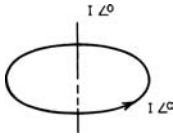


FIGURE 26-12 Horizontal loop and vertical dipole

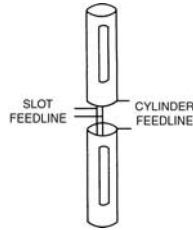


FIGURE 26-13 Slotted-cylinder circularly polarized antenna

pattern is circularly polarized at all points since the individual pattern of the loop and dipole are essentially the same. However, in practice this is difficult to obtain, except over narrow bandwidths, because of the different impedance characteristics of the loop and dipole.

A second combination is shown in Figure 26-13, consisting of two vertical half-wave-length-long cylinders in which vertical slots are cut.¹⁷ Feeding the two vertical cylinders will give a vertically polarized omnidirectional pattern in the plane normal to the axis of the cylinders, while feeding the two slots will give a horizontally polarized pattern in the same plane. If the power to both feeding arrangements is adjusted to be equal and the phase adjusted by controlling the length of the feedlines so that the two are in time-phase quadrature, the resulting pattern will be circularly polarized.

A combination of a slot and a dipole has been developed for a more directional antenna, e.g., as an element for a phased array.¹⁸ The configuration, which is shown schematically in Figure 26-14, uses a waveguide as the slot and a printed dipole as the complementary element. Since the waveguide and dipole radiation patterns track over a wide range of angles, this combined element can be used to generate any arbitrary polarization by controlling the complex excitations of the individual elements. Some overall degradation in the polarization performance will result from the difference in phase centers between the two individual elements.

The normal radiation (broadside) mode of a helix can also be considered as a combination of electric and magnetic antennas (dipoles and loops) producing circular polarization. This type of antenna has been discussed in Chapter 13.

A pair of crossed slots in the broad wall of a rectangular waveguide in which the field configurations in the waveguide are such that one slot is in time-phase quadrature with the other can also be considered as a combination of electric and magnetic antennas producing circular polarization. The combination will be discussed in greater detail shortly.

Two or more similar antennas when properly oriented in either time phase or space phase or a combination of body may be used to give a circularly polarized radiation field. A simple case is a pair of crossed half-wavelength dipoles. In Figure 26-15 circular polarization is obtained by having the equal currents in the dipoles in phase quadrature.

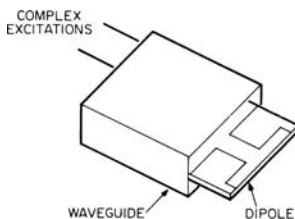


FIGURE 26-14 Slot-dipole element

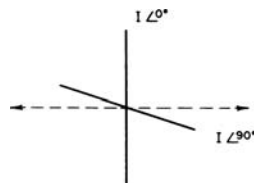


FIGURE 26-15 Crossed dipoles—current in phase quadrature

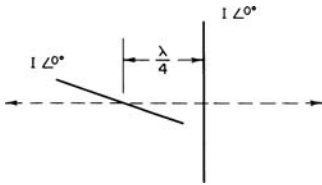


FIGURE 26-16 Crossed dipoles—current in phase— $\lambda/4$ separation

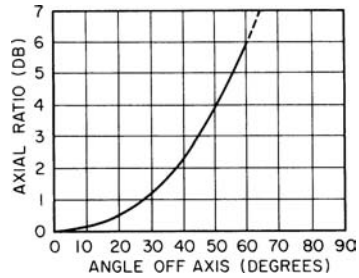


FIGURE 26-17 Deviation of circularity as a function of off-axis angle for a pair of crossed dipoles

Radiation is right-circularly polarized in one direction and left-circularly polarized in the opposite direction. If the pair of crossed dipoles is fed in phase and separated in space by a quarter wavelength as shown in Figure 26-16, circular polarization is again produced; however, the sense is the same in both directions. In both combinations the resulting field is circularly polarized only on axis. The deviation in circularity as a function of the off-axis angle is plotted in Figure 26-17.

Another case of a simple combination of similar antennas to produce circular polarization is a pair of narrow slots at right angles and located at the proper point in the broad wall of a rectangular waveguide.¹⁹ This may be explained by noting that the equations for the transverse and longitudinal magnetic fields of the dominant (TE_{10}) mode in a rectangular waveguide (see Figure 26-18) are

$$H_x = H_0 \sqrt{1 - \left(\frac{\lambda}{2a}\right)^2} \sin \frac{\pi x}{a} \quad (26-8)$$

$$H_z = -jH_0 \frac{\lambda}{2a} \cos \frac{\pi x}{a} \quad (26-9)$$

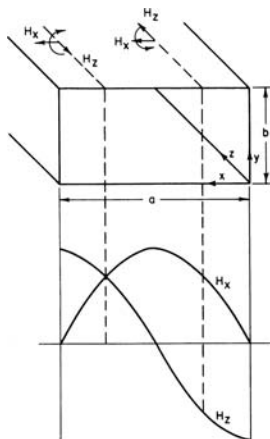


FIGURE 26-18 Field configuration, TE_{10} mode

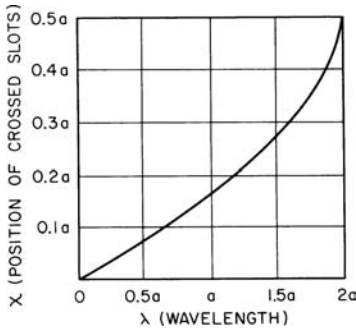


FIGURE 26-19 Position of crossed slots for circular polarization versus wavelength

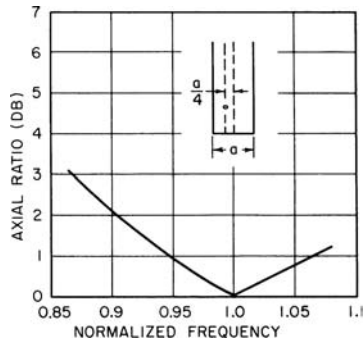


FIGURE 26-20 Theoretical axial ratio for $x = a/4$

From these two equations it may be seen that the fields are in phase quadrature and there are two values of x at which $|H_x| = |H_y|$. The values of x are given by

$$x = \frac{a}{\pi} \cot^{-1} \left[\pm \sqrt{\left(\frac{2a}{\lambda} \right)^2 - 1} \right] \quad (26-10)$$

Two crossed slots at either of these points will then radiate circularly polarized energy. Figure 26-19 is a plot of x versus λ over the wavelength range between the cutoff of the TE_{10} and TE_{20} modes. The orientation of the slots is arbitrary, and they may be made resonant and thus radiate a large amount of power (see Chapter 8). The theoretical axial ratio for $x = a/4$ is shown in Figure 26-20, which gives circular polarization at a frequency for which $\lambda = 2a/\sqrt{2}$.

26.4 DUAL-MODE HORN RADIATORS

A conventional waveguide horn may be used for the radiation and beaming of circularly polarized waves provided that it is fed with waveguide capable of propagating vertically and horizontally polarized waves simultaneously. The horn may be either symmetrical or asymmetrical, that is, square (round) or rectangular (elliptical). Figure 26-21 illustrates two types.

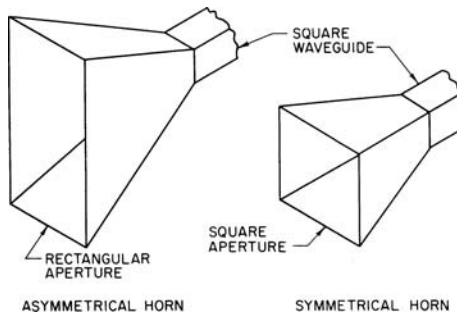


FIGURE 26-21 Examples of symmetrical and asymmetrical dual-mode horns

Symmetrical Case

A circularly polarized field will be obtained on the peak of the radiation pattern when the horn is fed through the square waveguide with equal-amplitude vertically and horizontally polarized modes arranged to be in quadrature. The radiated field will not, in general, be circularly polarized at other points on the radiation pattern because the vertically and horizontally polarized radiation patterns will have different beamwidths.

There are several methods useful in compensating for the two different beamwidths. Figure 26-22 shows two methods applied to the *azimuth plane* only of the horn. In actual practice, it is necessary to apply one of these methods (usually method *B*) to the elevation plane as well as to the azimuth plane.

When designing a circularly polarized horn for some specific radiation-pattern width, the standard design methods (such as those in Chapter 6 and Rhodes²⁰) are applicable, bearing in mind, of course, that an *E*-plane dimension for one polarization is an *H*-plane dimension for the other and also bearing in mind the effects of any compensating scheme.

When the horn is designed along conventional pyramidal or sectoral lines, phase tracking for the vertically and horizontally polarized fields is usually not a problem; that is, the radiation centers of phase for the two polarizations are sufficiently close so that differential phase-shift variations over the radiation pattern usually do not exceed about 10 to 15° within the tenth-power points on the beam.

Horns are sometimes constructed with a wide flange around the aperture for mechanical reasons. On small horns especially, the flange width should be no larger than necessary to avoid degradation of phase tracking.

Simple conical horns have beamwidths more alike for the two principal polarizations than do pyramidal horns. Controlled excitation of TM_{11} modes can further refine beamwidth matching. However, such structures are not as conveniently connected to square or rectangular waveguide as pyramidal horns.²¹

Corrugated conical horns offer excellent radiation-pattern amplitude and phase tracking for the two principal polarizations over bandwidths of half an octave³ or more.

Measurements reported in the literature²² show that cross-polarization components suppressed 30 dB or more from 8 to 11 GHz over all angles within approximately $\pm 45^\circ$ of boresight. The conical-horn geometry is shown in Figure 26-23. These structures tend to be larger for a given beamwidth than those of Figure 26-22 and may limit the closeness of adjacent secondary-beam pointing when they are used as juxtaposed feeds. Figure 26-24 shows the relationship between polarization ellipticity and cross-polarization component.

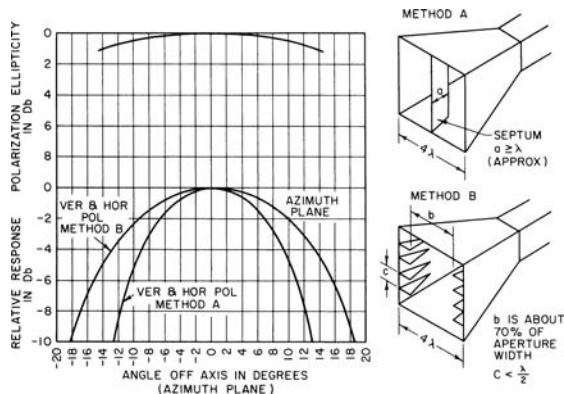


FIGURE 26-22 Two methods of compensating dual-mode horns for improved circularity

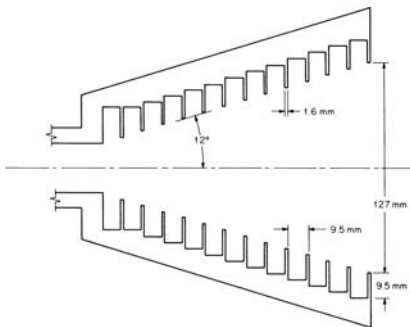


FIGURE 26-23 Section through conical corrugated horn—9-GHz nominal design frequency

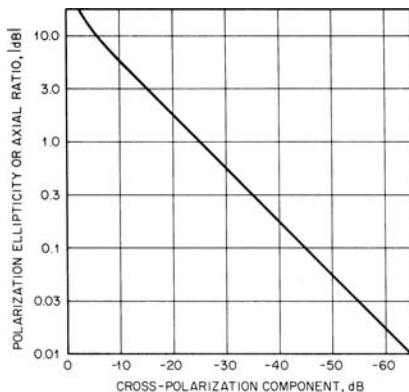


FIGURE 26-24 Relationship between polarization ellipticity or axial ratio and cross-polarization component

Asymmetrical Case

The preceding discussion is also applicable to asymmetrical horns, except that circularly polarized fields on the peak of the radiation pattern will not be obtained unless allowance is made for the difference in phase velocity of the vertically and horizontally polarized waves within the asymmetrical-horn flare. The differential phase shift may attain quite large values. For example, a measured differential phase shift of about 220° has been observed at 2800 MHz in a horn having a flare length of about 14 in (356 mm), a width of 2.84 in (72.1 mm), and a height of 7.8 in (198 mm).

The magnitude of differential phase shift can be computed to an accuracy of perhaps 10 percent by evaluating the integral

$$\text{Differential phase shift} = \int_0^L \frac{[\beta(\ell) - \beta(\ell)]}{V_{\text{pol}} H_{\text{pol}}} d\ell \quad (26-11)$$

in which L is the flared length of the horn and $\beta(\ell)$ is based simply on the appropriate width of the flare and the operating wavelength. Figure 26-25 shows the information in detail. Note that the differential phase shift may vary fairly rapidly with frequency. In the example mentioned earlier, the differential phase shift increased by about 15° for every 100-MHz decrease in frequency between 2900 and 2700 MHz.

Methods of Obtaining Quadrature

Quadrature phase relationship between two orthogonally polarized modes in round or square waveguide may be achieved by any of the following techniques:

1. Rectangular or elliptical cross section
2. Ridge guide
3. Dielectric slab
4. Multiple-lumped-element loading
5. Turnstile junction

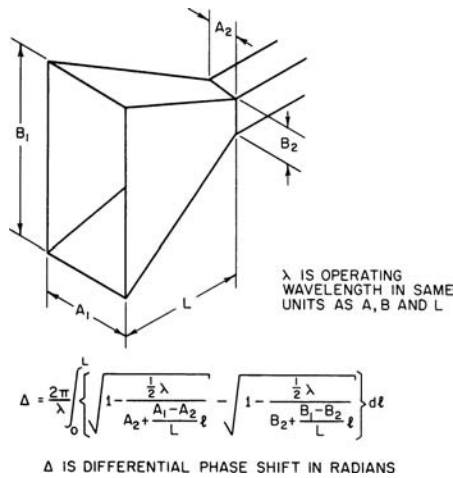


FIGURE 26-25 Method of determining differential phase shift in a sectoral horn

The choice of method must be based on the application, keeping such things in mind as ease of design, bandwidth, power-handling capacity, ease of adjustment, and ease of fabrication. Table 26-3 shows the trend of these characteristics in order.

The first three methods listed in the table are based on causing the phase velocity of the vertically and horizontally polarized modes to differ within the same section of waveguide. This section of waveguide will then be of different phase length for the two polarizations. Selection of the appropriate length to furnish the requisite differential phase shift is the last step in the design process.

For example, a rectangular section of waveguide carrying TE_{01} and TE_{10} modes and having internal dimensions a and b will produce differential phase shift at the rate of

$$\frac{2\pi}{\lambda} \left[\sqrt{1 - \left(\frac{\lambda}{2a}\right)^2} - \sqrt{1 - \left(\frac{\lambda}{2b}\right)^2} \right] \quad \text{rad/unit length} \quad (26-12)$$

TABLE 26-3 Preference Trends for Methods for Obtaining Phase Quadrature

Method	Ease of Design	Bandwidth	Power-handling Capacity	Ease of Adjustment	Ease of Fabrication (small quantity)
Rectangular or elliptical cross-section guide	1	5	1	5	4
Ridge guide	4	3	2	2	3
Dielectric slab	5	2	5	3	1
Multiple-lumped-element loading	2	1	3	1	2
Turnstile junction	3	4	4	4	5

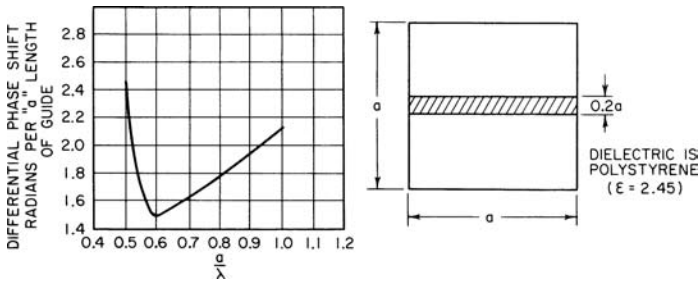


FIGURE 26-26 Differential phase shift in a partially loaded square waveguide

The values for a , b , and λ must be in the same units, of course. If such a section is to be connected to a square waveguide, it will be necessary to employ a suitable intermediate transformer section. Quarter-wavelength sections having an impedance equal to the geometric mean of the input and output impedances are satisfactory. Note that the differential phase shift (which can be computed with the aid of the preceding formula) of the transformer will contribute to the total.

Ridge waveguide can be handled in the same manner as the preceding rectangular guide. In the ridge case, however, the quantities $2a$ and $2b$ in the preceding expression should be replaced with λ_{CUTOFF} for TE_{01} and λ_{CUTOFF} for TE_{10} , respectively.

Dielectric-slab loading in a square waveguide will also produce a difference in phase velocity for the two modes of interest. Figure 26-26 provides details for a typical dielectric loaded waveguide. Just as earlier, it is necessary to provide means for obtaining an impedance match at each end of the phasing section. It is not sufficient as a rule to have the phasing section as a whole appear matched; it is important to have each end matched separately. If this precaution is not observed, it will be very difficult to predict total phase shift as well as the change in phase shift with a change in geometry. Consequently, adjustment or tailoring is difficult. Power-handling capacity will be somewhat lower in the loaded waveguide than in the unloaded case.

Lumped-element loading in a square waveguide has also been used successfully. The use of probe pairs is particularly convenient. Pair doublets and pair triplets are the most common, with the latter providing greater bandwidth and greater power-handling capacity for a given differential phase shift. Figure 26-27 shows design data for the doublet and triplet cases. The length of the probe necessary for a given value of susceptance is shown in Figure 26-28.

Circularly polarized waves can be generated directly in a waveguide junction

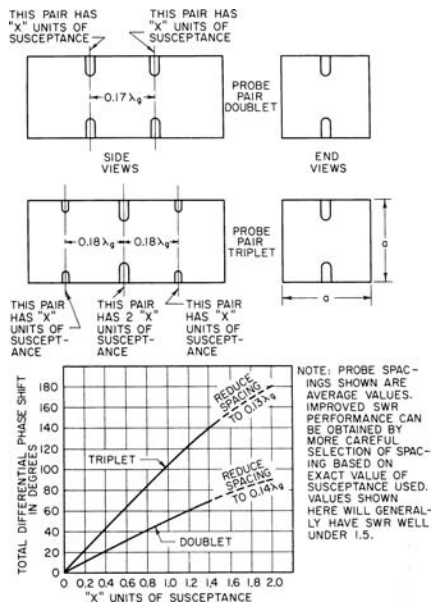


FIGURE 26-27 Probe phase-shifter design

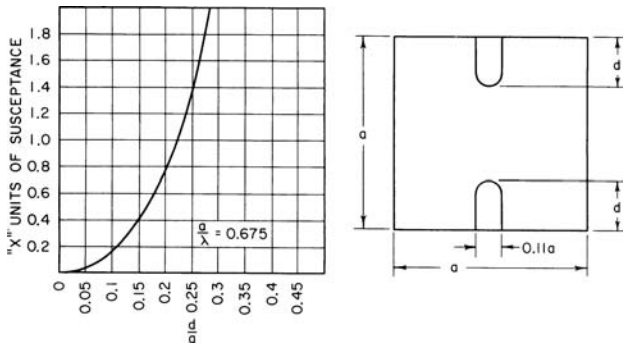


FIGURE 26-28 Susceptance as a function of probe penetration

known as a *turnstile junction* (not to be confused with a turnstile antenna). The general geometry is illustrated in Figure 26-29.

Dual-Mode Generation

The generation of two orthogonally polarized modes in a section of square or round waveguide can be accomplished in several ways, of which the two most common are

1. Coaxial cable probes at right angles
2. Conventional rectangular waveguide inclined at 45°

The second case is particularly convenient because the ratio of magnitudes of the vertically and horizontally polarized waves can be altered merely by altering the inclination of the rectangular guide.

In generating and guiding dual modes in square or round guide, care should be employed to keep the geometry symmetrical about each of the two principal longitudinal planes. Any asymmetry will tend to excite the TM_{11} or TE_{11} modes. For square or round guide, these two modes will usually not be very far from cutoff. For example, in square guide, with dimensions a high and a wide, cutoff for the TE_{01} and TE_{10} modes occurs at $\lambda = 2a$. The cutoff for TM_{11} and TE_{11} modes occurs at $\lambda = 2\sqrt{2}a$, or only about 41 percent above cutoff for the dominant modes.

In all the discussion earlier concerning phase shift, it has been assumed that there is an impedance match for each of the generated dual modes. The existence of mismatch may produce an unexpected alteration in the phase shift. For example, if there is a perfect match

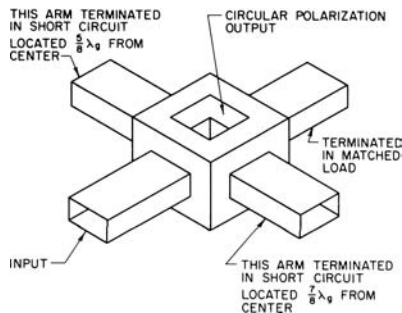


FIGURE 26-29 Turnstile junction

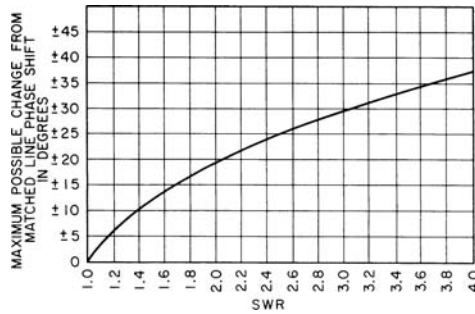


FIGURE 26-30 Possible phase variation with standing-wave ratio

for both the vertically and the horizontally polarized modes and the differential phase shift noted, the appearance of a mismatch in one of the modes may produce a change in differential phase shift, as shown in Figure 26-30. If the same magnitude of mismatch appears in both modes, the change may be as large as twice that of Figure 26-30.

26.5 MICROSTRIP PATCH RADIATORS

Circular polarization can be produced by microstrip patch antennas in a variety of manners. The methods for producing CP in a patch antenna are often described as either single or dual feed types. Dual feed patches are normally fed at two points on the antenna with equal amplitude and 90° out of phase using either a hybrid power divider or a quarter-wave offset feed.²³ Dual feed methods have the advantage of being conceptually easier to design even though they require external components. The single feed method must excite two orthogonal modes with one feed and without the use of external hybrids. This is accomplished by carefully choosing the radiating patch geometry such that orthogonal degenerate modes are properly excited to produce circular polarization. These *degenerate* modes have equal or nearly equal amplitudes and radiate in directions orthogonal to each other.

Figure 26-31 shows three common methods of creating circular polarization in a square or nearly square patch with a single feed. The antenna in Figure 26-31a is an asymmetric or nearly square patch²⁴ that has one edge length slightly greater than the other, $\Delta/a \ll 1$. Feeding the antenna at either point 1 or 3 will excite the mode normal to that particular edge,

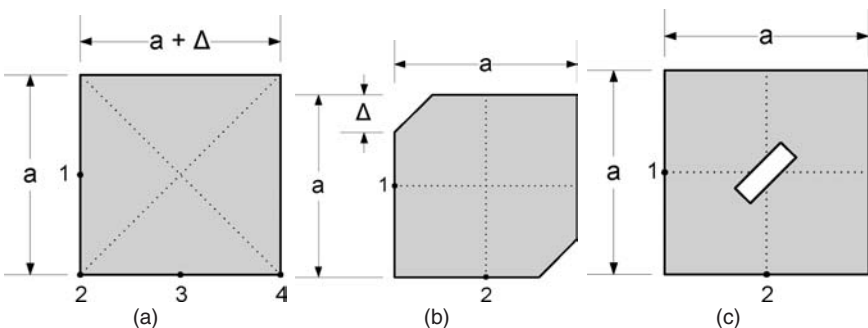


FIGURE 26-31 Single feed microstrip patches for circular polarization. Dashed lines represent single feed locations capable of producing circular polarization.

and the antenna will behave as a linearly polarized microstrip patch antenna. However, if the antenna is fed from either point 2 or 4, or along either patch bisecting diagonal, two orthogonal modes will be excited and circular polarization will be created. It is possible to produce either right- or left-hand sense circular polarization depending on the selection of feed location. The position of the feed along the diagonal also has a significant impact on the input impedance seen by the patch.

The antenna in Figure 26-31*b* is a corner-truncated square patch. Two opposing corners are trimmed a small amount, Δ , where $\Delta/a \sim 0.1$. This antenna creates circular polarization in a manner similar to the nearly square patch of Figure 26-31*a* by creating two orthogonally degenerate modes from the slight perturbation in antenna geometry due to the truncated corners.²⁵ In this case, however, the antenna must be fed from point 1 or 3, or from a point along either of the antenna center lines. Both right- and left-hand sense circular polarization may be established with this geometry depending on which feedpoint is chosen. If the antenna is fed along the diagonals, only linear polarization will be produced. The antenna in Figure 26-31*c* is a slotted patch, and it also produces circular polarization by creating a small perturbation in the geometric symmetry. This antenna has a diagonal slot cut in the radiating patch that is angled at 45° from the patch edge. The length and width of the slot are design dependent. Like the truncated square patch, the slotted patch must be fed from either point 1 or 3, or from either of the lines that are diagonal to the patch geometry perturbation to produce circular polarization along. Additional details on the subject of microstrip patch antennas can be found in Chapter 7.

26.6 TRANSMISSION- AND REFLECTION-TYPE POLARIZERS

A transmission polarizer is an anisotropic propagation medium whose anisotropy is adjusted to achieve time (i.e., phase) quadrature for two waves whose (linear) polarization vectors are mutually orthogonal. Such polarizers operate independently of the nature of the source of the waves, are often relatively bulky and awkward to adjust, can be very inexpensive to manufacture, and can have very high power-handling capability.

The simplest type of polarizer is a parallel-metal-plate structure (see Figure 26-32). In this structure, the incident linearly polarized energy is inclined 45° to the metal-plate edges, so that there are two equal field components, one parallel to the plates and one perpendicular to the plates. The component perpendicular to the plates passes through the structure relatively undisturbed, while the component parallel to the plates experiences a phase shift relative to free-space propagation. If the spacing and length of the plates are adjusted so that the field parallel to the plates advances $\lambda/4$ with respect to the field perpendicular to the plates, the two fields at the exit of the plates result in a circularly polarized wave. This structure is commonly referred to as a *quarter-wave plate*.

Any anisotropic dielectric can be used as a transmission-type polarizer, provided only that

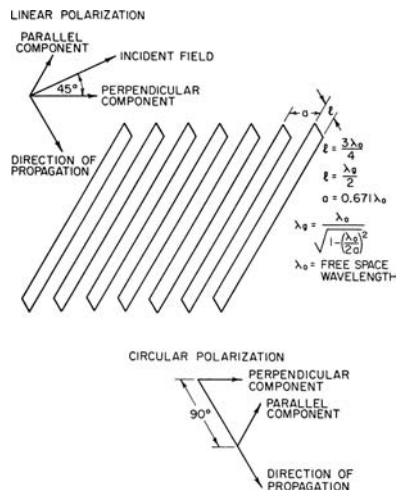


FIGURE 26-32 Parallel-plate polarizer

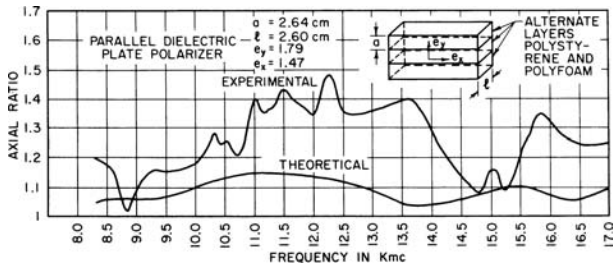


FIGURE 26-33 Parallel-dielectric-plate polarizer

positioning and properties result in two equal-amplitude, quadrature-phased linearly polarized waves at the exit port.

There are several ways to produce such anisotropy: (1) parallel metal plates,^{26,27} (2) parallel dielectric plates²⁸ (see Figure 26-33), and (3) a lattice structure composed of strips or rods.^{29,30} The first two methods are the heaviest, while the last two are the least frequency-sensitive. An unfortunate complication in design and adjustment of the simplest configurations is caused by imperfect impedance matching at the entrance and exit ports; in general, it is different for the two principal polarizations. Unwanted amplitude and phase effects result, including the result of a nonunity standing-wave ratio (SWR) inside the polarizer.

A variation of the Method 3 just mentioned is the meander-line polarizer.³¹⁻³⁴ Figure 26-34a shows a typical meandering configuration of a thin metal conductor supported on a dielectric sheet. In general, a greater number of layers of such sheets yields greater bandwidths, just as in multiple discontinuities used for broadband transmission-line transformers.

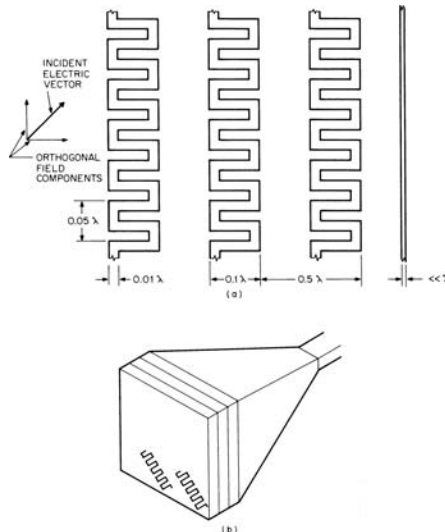


FIGURE 26-34 Typical meander-line dimensions and radome applications: (a) Typical nominal dimensions of conductors on a layer (not to scale); (b) Four-layer polarizer-radome on a horn, illustrating orientation angle of typical conductors (not to scale)

Kotlarenko³¹ reports bandwidths of 1.7 to 1 for axial ratios of less than 3 dB for incidence angles ranging over 30° when using a six-layer design. A five-layer design used at normal incidence yielded axial ratios under 3 dB for bandwidth exceeding an octave.

Epis³⁴ reports axial ratios under 3 dB over bandwidths of 1.75 to 1 by using a four-layer meander-line structure as a horn radome and a bandwidth exceeding 2.5 to 1 with axial ratios below 1.8 dB with six layers.

Reflection-type polarizers are essentially half-length transmission-type polarizers mounted on a conducting sheet. The simplest type of reflection-type polarizer is a series of closely spaced metal vanes, one-eighth wavelength high, on a conducting sheet. The incident energy is polarized at 45° to the vanes, so that the component parallel to the vanes is reflected by edges of the vanes and the component perpendicular to the vanes is reflected by the conducting sheet, thus delaying it 90° with respect to the parallel component and producing circular polarization. In general, any of the anisotropic propagating systems described earlier may be halved longitudinally and, with proper design, mounted above a conducting sheet to form a reflection polarizer.

Circular polarization may also be obtained via reflection from a conventional lossless dielectric when the dielectric constant ϵ exceeds 5.8 and the incidence angle θ satisfies the condition $\sin^2 \theta = 2/(\epsilon + 1)$.³⁵

26.7 POLARIZATION MEASUREMENTS

The measurement of circularly polarized antennas or antennas of unknown polarization is generally similar to that of other antenna types, except that it is often necessary to measure the state of polarization (degree of polarization ellipticity, or axial ratio) as well as the other more customary parameters. Several methods are commonly used to measure the additional parameters required to characterize antenna polarization. These methods include the radiation pattern methods, polarization pattern methods, amplitude-phase methods, and multiple component methods.³⁶

The radiation pattern method is convenient and is generally performed using either a single linearly polarized probe antenna or two orthogonally polarized antennas. In the single antenna variation, the probe antenna is arranged to rotate about its boresight axis at an angular velocity, ω_p , as the unknown test antenna is rotated through its desired pattern cut at an angular velocity, ω_θ . Provided that the rotational velocity of the probe is much quicker than the rotational velocity of the test antenna ($\omega_p \gg \omega_\theta$), the resulting pattern will give the ellipticity magnitude as a function of the pattern angle. This method, however, will not give the right-hand or left-hand sense, which must be separately determined.

Attention is called here to the unusual degree of sensitivity of polarization measurement to small distorting influences of secondary paths in a measuring setup. Figure 26-35 illustrates the conditions.

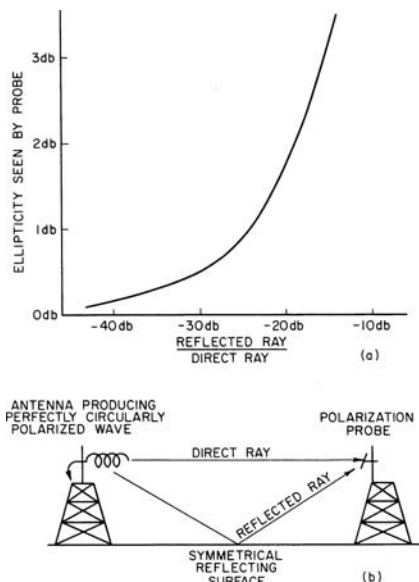


FIGURE 26-35 Influence of the reflected wave on the measurement of a circularly polarized wave

REFERENCES

1. *IEEE Standard Test Procedures for Antennas*, IEEE Std. 149-1979, Institute of Electrical and Electronics Engineers, 345 East 47th Street, New York, NY 10017.
2. W. B. Offutt, "A Review of Circular Polarization as a Means of Precipitation-Clutter Suppression and Examples," *Nat. Electron. Conf. Proc.*, vol. 11, 1955.
3. R. W. Kreutel, D. F. DiFonzo, W. J. English, and R. W. Gruner, "Antenna Technology for Frequency Re-Use Satellite Communications," *IEEE Proc.*, vol. 65, no. 3 (March 1977). A comprehensive survey paper containing a good bibliography.
4. W. L. Stutzman and W. P. Overstreet, "Axial Radio Measurements of Dual Circularly Polarized Antennas," *Microwave J.*, vol. 24, no. 10 (October 1981): 49-57. A method for polarization properties of antennas installed in operating position.
5. W. Sichak and S. Milazzo, "Antennas for Circular Polarization," *IRE Proc.*, vol. 36, no. 8 (August 1948): 997-1001. Derivation of expression for voltage induced in an arbitrarily polarized antenna by an arbitrarily polarized wave, plus typical applications.
6. L. Hatkin, "Elliptically Polarized Waves," *IRE Proc.*, vol. 38, no. 12 (December 1950): 1455. Simplified derivation of expression for voltage induced in an arbitrarily polarized antenna by an arbitrarily polarized wave.
7. A. C. Ludwig, "A Simple Graph for Determining Polarization Loss," *Microwave J.*, vol. 19, no. 9 (September 1976): 63.
8. W. L. Flock, *Propagation Effects on Satellite Systems at Frequencies Below 10 GHz, A Handbook for Satellite Systems Design* (Washington, DC: NASA Reference Publication 1108(02), 1987).
9. I. Koffman, "Feed Polarization for Parallel Currents in Reflectors Generated by Conic Sections," *IEEE Trans. Antennas Propagat.*, vol. AP-13, no. 1 (January 1966): 37-40. Presentation of a method for determining the ideal feed polarization that should be incident on a conic reflector.
10. P. A. Watson and S. I. Ghobrial, "Off-Axis Polarization Characteristics of Cassegrainian and Front-Fed Paraboloidal Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-20, no. 6 (November 1972): 691-698. Comparison of the amount of cross-polarized energy for Cassegrainian and front-end paraboloids.
11. A. C. Ludwig, "The Definition of Cross Polarization," *IEEE Trans. Antennas Propagat.*, vol. AP-21, no. 1 (January 1973): 116-119. Presentation of a discussion of three definitions of cross-polarization with respect to several applications. It is noted that some confusion can exist depending on the definition.
12. T. S. Chu and R. H. Turrin, "Depolarization Properties of Offset Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-21, no. 3 (May 1973): 339-345. Presentation of results for both linearly and circularly polarized feeds. Measurements verify the analyses.
13. J. Jacobsen, "On the Cross Polarization of Asymmetric Reflector Antennas for Satellite Applications," *IEEE Trans. Antennas Propagat.*, vol. AP-25, no. 2 (March 1977): 276-283. Suggestions on how to design low cross-polarized feeds for offset-fed antennas.
14. Y. Rahmati-Samii and A. B. Salmasi, "Vectorial and Scalar Approaches for Determination of Interbeam Isolation of Multiple Beam Antennas—A Comparative Study," *Antennas Propagat. Int. Symp. Dig.*, vol. 1 (1981): 135-139. Discussion of three approaches to calculating interbeam isolation and frequency-reuse characteristics of multiple-beam antennas.
15. A. C. Ludwig, "Antenna Feed Efficiency," *Space Programs Summary 37-26*, Jet Propulsion Laboratory, California Institute of Technology (1965): 200-208. Discussion of the factors relating to feed efficiency.
16. "30/20 GHz Spacecraft Multiple Beam Antenna System Concept for Demonstration Satellite," Cont. NAS 3-22499, TRW Defense and Space Systems Group, April 22, 1981. Description of the design parameters of offset multiple-beam antennas.
17. C. E. Smith and R. A. Fouty, "Circular Polarization in F-M Broadcasting," *Electronics*, vol. 21 (September 1948): 103-107. Application of the slotted cylinder for a circularly polarized omnidirectional antenna.

18. R. M. Cox and W. E. Rupp, "Circularly Polarized Phased Array Antenna Element," *IEEE Trans. Antennas Propagat.*, vol. AP-18, no. 6 (November 1970): 804–807. Description and measurements of a combined waveguide and dipole element.
19. A. J. Simmons, "Circularly Polarized Slot Radiators," *IRE Trans. Antennas Propagat.*, vol. AP-5, no. 1 (January 1957): 31–36. Theoretical and experimental investigation of producing circular polarization by using two crossed slots in the broad wall of a rectangular waveguide.
20. D. R. Rhodes, "An Experimental Investigation of the Radiation Patterns of Electromagnetic Horn Antennas," *IRE Proc.*, vol. 36, no. 9 (September 1948): 1101–1105. Presentation containing design data for horn radiators.
21. P. D. Potter, "A New Horn Antenna with Suppressed Sidelobes and Equal Beamwidths," *Microwave J.*, vol. 6, no. 6 (June 1963): 71–78. Theory and design of a circularly symmetrical horn having steps to control the TE_{11} mode advantageously.
22. P. J. B. Clarricoats, A. D. Olver, and C. G. Parini, "Optimum Design of Corrugated Feeds for Low Cross-Polar Radiation," *Proc. Sixth European Microwave Conf.* (1976): 148–152.
23. J. R. James and P. S. Hall (eds.), *Handbook of Microstrip Antennas*, vol. 28 of IEE Electromagnetic Waves Series (London: Peter Peregrinus Ltd., 1989).
24. W. F. Richards, Y. T. Lo, and D. D. Harrison, "An Improved Theory for Microstrip Antennas and Applications," *IEEE Trans. Antennas Propagat.*, vol. 29 (January 1981): 38–46.
25. R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook* (Boston: Artech House, 2001).
26. J. Ruze, "Wide-Angle Metal Plate Optics," *IRE Proc.*, vol. 38, no. 1 (January 1950): 53–59. Derivation of the design equations for constrained metal-plate lenses.
27. J. F. Ramsay, "Circular Polarization for C. W. Radar," Marconi's Wireless Telegraph Co., Ltd., 1952. Proceedings of a conference on centimetric aerials for marine navigational radar held on June 15–16, 1950, in London.
28. H. S. Kirschbaum and L. Chen, "A Method of Producing Broadband Circular Polarization Employing an Anisotropic Dielectric," *IRE Trans. Microwave Theory Tech.*, vol. MTT-5, no. 3 (July 1957): 199–203. Description of a procedure whereby it is possible to design circular polarizers for both waveguide and window form to be used over a broad band of frequencies.
29. J. A. Brown, "The Design of Metallic Delay Dielectrics," *IEE Proc. (London)*, part III, vol. 97, no. 45 (January 1950): 45. Development for the simplest case of a theory of metallic delay dielectrics based on an analogy with shunt-loaded transmission lines when the delay medium consists of an array of infinitely long conducting strips.
30. H. S. Bennett, "The Electromagnetic Transmission Characteristics of the Two-Dimensional Lattice Medium," *J. App. Phys.*, vol. 24, no. 6 (June 1953): 785.
31. I. Kotlarenko, "Analysis and Synthesis of Electromagnetic Wave Polarizers," *Proc. Eleventh IEEE Conf.*, Israel, October 23–25, 1979. Theoretical discussion and practical results of linear conductors and meander-line conductors in multilayer-array polarizers.
32. T. L. Blakney, J. R. Burnett, and S. B. Cohn, "A Design Method for Meander Line Circular Polarizers," *22d Ann. Symp. USAF Antenna R&D Prog.*, University of Illinois, Urbana (October 1972): 1–15.
33. H. A. Burger, "A Dual Polarized Antenna System Using a Meander Line Polarizer," *Antennas Propagat. Int. Symp.* (May 1978): 55–58.
34. J. J. Epis, "Theoretical Analysis of Meander Line Array Radome Polarizers," Western Div., GTE Sylvania Electronic Systems Group, December 1972. Theoretical establishment of the complex impedance discontinuities associated with meander-line arrays.
35. J. A. Stratton, *Electromagnetic Theory* (New York: McGraw-Hill, 1941): 279–280, 499–500. Mathematical definition of an elliptically polarized wave and discussion of formation of a circularly polarized wave by total reflection from a dielectric interface.
36. W. L. Stutzman, *Polarization in Electromagnetic Systems* (Boston: Artech House, 1992): 72–82. Detailed description of polarization measurement methods.

P · A · R · T · 3

Applications

Chapter 27

Low- and Medium-Frequency Antennas

Nikolaos K. Uzunoglu

National Technical University of Athens

CONTENTS

27.1	INTRODUCTION	27-4
27.2	FUNDAMENTALS OF VERTICAL MONOPOLE ANTENNA RADIATION PROPERTIES	27-4
27.3	PRACTICAL LOW- AND MEDIUM-ANTENNA STRUCTURES: DESIGN AND CONSTRUCTION BASICS	27-7
27.4	NUMERICAL COMPUTATIONS OF LOW-FREQUENCY ANTENNAS	27-19
27.5	VERTICAL MAGNETIC LOOP AND HORIZONTAL SLOT ELF AND VLF ANTENNAS USING NATURAL STRUCTURES	27-20

27.1 INTRODUCTION

Following the experimental validation of James Maxwell's electromagnetic theory in 1884 by Heinrich Hertz, the first real technological application using electromagnetic waves in transmitting information signals was carried out by Guglielmo Marconi at the beginning of the 20th century.¹ Marconi realized from the beginning of his work that the use of vertical metallic towers as antennas enables the radiation of low-frequency (LF) waves in the region of 50–100-kilohertz (kHz) and also provides efficient propagation by the ground wave propagation mechanism, although the exact computation of the field in this mode went through much testing, starting from the pioneering work of Sommerfeld and lasting into the 1940s with the work of Norton.² During this early period of radio communications, it was also realized that vertical mounted antennas could support long-distance ionosphere communication links of single or multiple hops. Because of this, in the early period of radio communication systems the broadcasting and point-to-point links mostly used vertical monopole antennas constructed by using vertical metallic towers and utilizing the low- (30–300-kHz) and medium- (300–3000-kHz) frequency bands. Despite the growing use of higher frequencies—starting in the 1930s with the high (3–30 MHz) and later very high (30–300 MHz), ultra high (300–3000 MHz), and microwave (>3 GHz) frequencies—the importance of low and medium frequencies has not diminished, because they offer unique properties such as very stable propagation conditions and the ability to penetrate the sea and earth.

A fundamental difficulty in achieving efficient radiation at low frequencies is the necessity of having antenna dimensions comparable to radiation wavelength (λ) in the air. Restricting the antenna geometry to a vertical monopole of height h , the radiation efficiency drops with the ratio $(h/\lambda)^2$ as the wavelength increases. Maximum efficiency is obtained if the antenna height is $h = \lambda/4$, also providing that a good match to a low-impedance transmission line (50- Ω) is achieved. For medium frequencies (MF) at 600 kHz at the radiation wavelength $\lambda = 500$ m, maximum efficiency is achieved if $h = 125$ m, which doesn't pose a serious construction problem. However, at the low edge of low frequencies with a corresponding height of much more than 500 m, you have to restrict the antenna height to less than 400 m. The basic environmental condition restricting the antenna height is the wind load. In particular, wind loads exhibiting no homogeneity along the vertical axis could pose serious challenges to the mechanical stability of the mast. Earthquake-induced forces are not significant.

Although the possible radiation mechanism at low frequencies is not restricted only to a vertical monopole, as will be discussed later in Section 27.5, vertical monopole antennas predominately are used in constructing very-low- (10–30-kHz) and low-frequency (30–300-kHz) radiation systems.

Another unique property of low-frequency antennas is related to the insignificance of the receiving antenna's dimensions. This is because at low frequencies atmospheric noise exceeds the receiver internal noise, and therefore the receiver output signal-to-noise power ratio is independent of antenna efficiency and size.

27.2 FUNDAMENTALS OF VERTICAL MONOPOLE ANTENNA RADIATION PROPERTIES

The radiation properties of a vertical monopole antenna, shown in Figure 27-1 and having height h , are easily computed by using the image theory, assuming the earth surface is a perfect conductor. For an earth medium with good conductivity $\sigma > 0.01$ S/m (Siemens/meter),

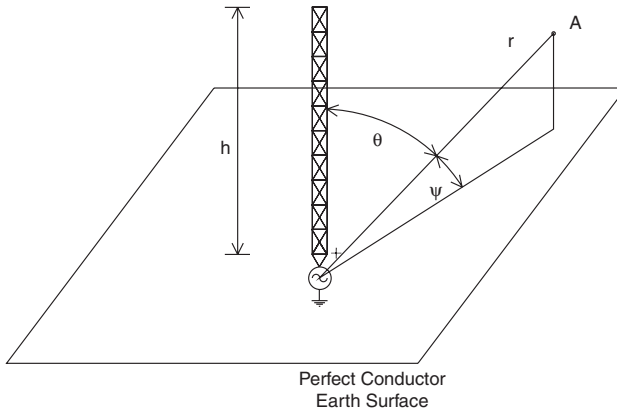


FIGURE 27-1 Vertical monopole antenna geometry

in practical terms, the ground surface can be assumed to have perfect conductivity. Usually wire grids (see “VLF and LF Antennas” in Section 27.3) are used to achieve good conductivity in the vicinity of the antenna. Therefore, to understand the fundamental radiation properties of low-frequency vertical monopole antennas, the earth surface could be assumed to be a perfect conductor. Then, based on fundamental concepts of antenna theory, the total electric current distribution on the vertical antenna tower is almost equal to a sinusoidal function as

$$I(z) \approx I_0 \sin(k_0(h-z)) / \sin(k_0 h) \quad (27-1)$$

where z is the distance measured from the base of the monopole, I_0 is the base current, and $k_0 = 2\pi/\lambda$ is the free space propagation constant. At very low frequencies (VLF), $h \ll \lambda$, Eq. 27-1 reduces to a linear relationship, $I(z) \approx I_0 k_0 (h-z)/h$. The antenna characteristics are defined in terms of its radiation resistance

$$R_r = 160\pi^2 (h_e/\lambda)^2 \quad (27-2)$$

where h_e is the equivalent antenna height computed by the equation

$$h_e = \frac{1}{I_0} \int_0^h dz I(z) \quad (27-3)$$

The antenna radiation efficiency is computed as the ratio of the radiated power to total power fed to the antenna:

$$\eta = \frac{R_r}{R_r + R_l} \quad (27-4)$$

where R_l is the total loss antenna resistance representing the non-radiated power consumed by the antenna structure and the ground medium in the vicinity of the monopole.

The radiation pattern that is identical to a Hertzian dipole is defined in terms of radiated power density (W/m^2) in free space:

$$P(\theta) = P(\pi/2) \sin^2(\theta) \quad (27-5)$$

where the radiation power density at the horizontal plane is

$$P(\pi/2) = \frac{22.75k_0^2 h_e^2 I_0^2}{r^2}$$

where $r \gg \lambda$ is the distance of field measurement point and θ is the angle between the vertical axis and the position vector.

From the early period of radio engineering (1910–1920), it was understood that the efficiency of a monopole antenna can be increased significantly by making the antenna current as constant as possible along the vertical axis. This is achieved by loading the antenna top with various conductor structures, as shown in Figure 27-2. Usually, a symmetric radial network of wire conductors as shown in Figure 27-2c is used. The loading increases the equivalent antenna height and therefore the radiation efficiency.

Top-loading increases the equivalent antenna height, h_e , and thus the radiation resistance and radiation efficiency. The antenna loading is accounted for by the existing capacitance, C_0 , between the top-loading and ground. The equivalent antenna height, h_e , can be computed by using Eq. 27-3 and the current distribution:

$$I(z) \approx I_0 \sin(k_0(h + s - z)) / \sin(k_0(h + s)) \quad (27-6)$$

where s is the slant length of the top-loading wires.

In modeling the low-frequency antenna, the equivalent circuit shown in Figure 27-3 is used for the input impedance of the monopole.

Based on the equivalent circuit, the input impedance of the antenna can be computed as follows:

$$Z_{in} = R_T(\omega) + j\omega L - 1/j\omega C = R_T(\omega) + jX_T(\omega) \quad (27-7)$$

where $R_T = R_r + R_l$ is the total antenna base resistance, L is the antenna mast inductance, and $C = C_a + C_0$ is the total (mast plus the top-loading wires) antenna capacitance. The reactance (imaginary) part of the impedance can be written in terms of the antenna resonance frequency

$$\omega_0 = 1/\sqrt{LC} \quad (27-8)$$

as follows:

$$X_T(\omega) = \omega L(1 - (\omega_0 / \omega)^2) \quad (27-9)$$

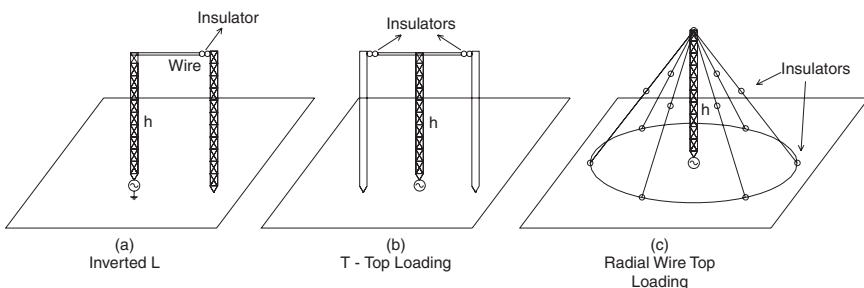


FIGURE 27-2 Three types of top-loading monopole antennas: (a) inverted L, (b) T top-loading, and (c) radial wire top-loading

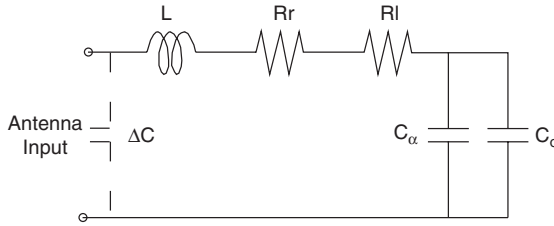


FIGURE 27-3 Monopole antenna equivalent circuit: L = antenna inductance, C_0 = top-loading capacitance, C_α = antenna mast capacitance, R_r = radiation resistance, R_l = total loss resistance, and ΔC = antenna feedpoint leak capacitance

Usually at VLF (3–30 kHz) the operation frequency is much less than the resonance frequency.

A critical parameter, especially at frequencies less than 100 kHz, is the antenna bandwidth determining the available communication channel capacity or the maximum data bit rate of transmission.³ The intrinsic 3-dB antenna bandwidth is computed³ using the definition of the quality (Q) parameter of the equivalent circuit of Figure 27-3:

$$BW_0 = f/Q = 1.102458 \cdot 10^{-7} Ch_e^2 f^4 \quad (27-10)$$

where $f = \omega/(2\pi)$ is the operation frequency. The overall bandwidth of the system is determined taking into account the antenna losses and is

$$BW_l = BW_0 / \eta \quad (27-11)$$

Above 150 kHz the antenna bandwidth restriction is limited, and traditionally, monopole antennas are used to broadcast double-side amplitude modulation signals.

The most common broadcast antenna type used in the frequency region of VLF, LF, and MF is the vertical monopole. However, in theory other types of antenna, especially vertical loop antennas, have been suggested to cope with the difficulty of constructing very tall masts needed at the low edge of the spectrum. This topic is analyzed in Section 27.4.

27.3 PRACTICAL LOW- AND MEDIUM-ANTENNA STRUCTURES: DESIGN AND CONSTRUCTION BASICS

Design and construction of low-frequency antennas poses formidable civil engineering tasks primarily having to do with the erection of the metallic tower structures, which could be as high as 400 m. Concerning the electromagnetic properties of the antenna, issues to be addressed are

- The conductivity properties of the ground environment where the antenna is to be installed and, in case of poor conductivity ($<10^{-2} S/m$), mounting the antenna inside the earth grid of wires
- Metallic mast, supporting steel ropes, and top-loading wires as well as high-voltage insulators to be used in connecting structures of different voltages and the antenna base insulator
- Antenna matching unit to achieve tuning of the antenna capacitive impedance to low-impedance (50- Ω) power amplifier output

Next, these issues are addressed for VLF, LF, and MF antennas.

VLF and LF Antennas

The most common antenna structure used at the 16-200-kHz band is the base-insulated top-loaded antenna already shown in Figure 27-2c. Alternative structures, as shown in Figure 27-2a and b, have also been used in the past, such as *T*- and inverted-*L*-shaped antennas. A spiral top-loading structure also has been proposed to achieve relatively small size, low feed voltage, and even to reduce the tuning circuit inductor size.⁴ Increased capacitance for VLF umbrella radial-wire top-loading antennas using multiple-wire rib construction also has been analyzed.⁵ The use of this type of top-loading wiring provides an approximately 20 percent increase of capacitance without any serious change of effective height of the monopole antenna. Other types of structure such as “Tratic” and “Trideco” antennas, shown in Figure 27-4, have been used in transmitting high power levels. The capability of lowering top-load wires without appreciably losing efficiency is an important property of this type of multiwire antenna.

The most commonly used structure is the base-fed monopole antenna using an umbrella type top-loading structure. Figure 27-5 shows various parts of this type of antenna. A site-assembled triangular tower of up to 600 m is used to construct the monopole. A highly critical component is the base ceramic insulation, which should provide sufficient high-voltage strength. In practical terms the base voltage is limited to 250 kilovolts (kV), thus restricting the highest power to be transmitted by each monopole antenna. The top-loading consists of 16–24 steel wires of 1–1.5-inch diameter connected directly to the tower’s top point, and then the active wires are discontinued by an insulator and continue with supporting guy wires, which are anchored to the ground through a second insulator to reduce the induction of loop currents due to capacitive coupling to supporting guy wires. The angle between the tower axis and active top-load wires should be larger than 50–60° to reduce the self-shielding phenomenon of the top-load wires. In addition to top-loading insulated guy wires, the tower should be supported by guy wires to the ground with a 120° angle separation at two or three levels along the tower height to provide sufficient resistance to wind load. For the supporting wire, guy insulators should be placed both at the tower and the ground anchoring point. Assuming the monopole height is h , the space to be allocated for the antenna installation should have a diameter of at least $3.5h$.

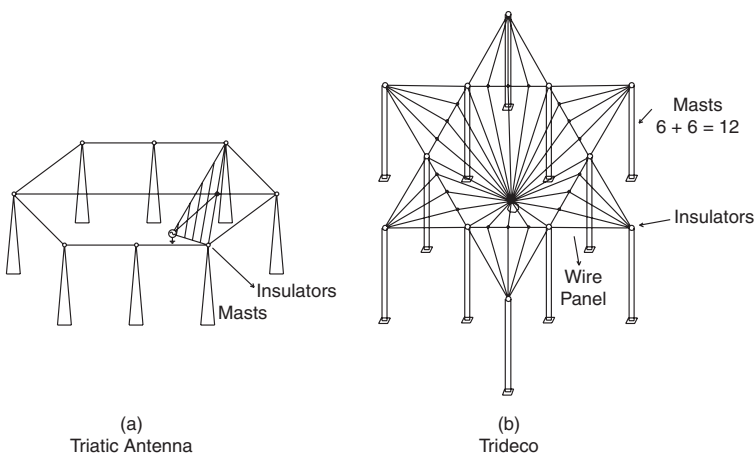


FIGURE 27-4 Tratic and Trideco VLF antennas

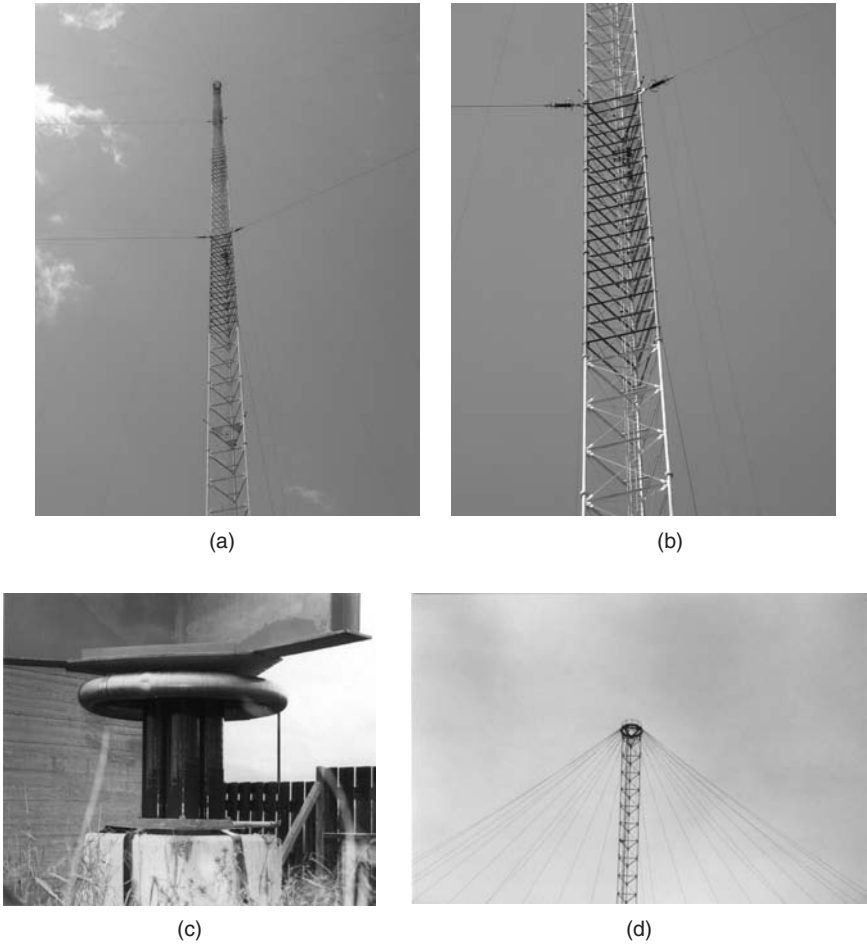


FIGURE 27-5 Umbrella-type top-loading VLF antenna: (a) LF antenna mast, (b) LF antenna supporting guys, (c) LF antenna base insulator-feeding point, and (d) LF top-loading

In Figure 27-6 the variation of top-load capacitance with the number of wires is shown for an antenna of $h = 250$ m height and top radial wires geometry shown before and depicted also in Figure 27-6. The computation was carried out by a method of moments (see Section 27.4).

Conductors Used in VLF-LF Antennas

The choice of conductor wires and structural cables is a critical issue. Top-loading conductors should have as low resistance as possible, taking into account the skin effects to guarantee low radio frequency (RF) losses. Furthermore, active wires should have adequate tensile strength and minimum weight. The conductor diameter and total conductor area

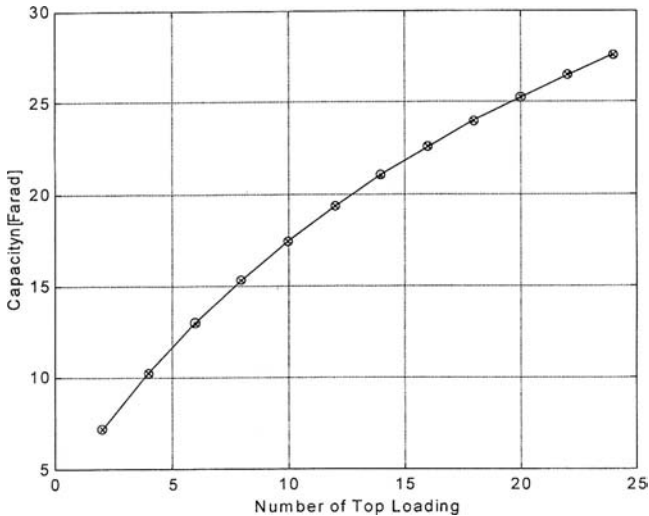


FIGURE 27-6 Variation of top-loading capacitance for an umbrella-type radial-wire conductor

must be sufficient to limit the surface potential to prevent a corona. A method developed by Wheeler⁶ provides a solution based on the average current flowing through the wires toward the ground per unit wire surface area. By using this method, the average voltage gradient on the conductor surface is

$$E_c \text{ (kV/mm)} = \frac{3\lambda^2}{2\pi h_e A_c} \sqrt{10P \text{ (watts)}} \times 10^{-6}$$

where E_c is the electric field intensity measured in kV/mm, A_c is the total wire surface in m^2 , and h_e and λ are measured in meters.

Electric power transmission cables such as concentric-layer, aluminum-clad steel or aluminum wire, aluminum-clad wire core, aluminum conductor, steel-reinforced type ropes could be used. The cables that have proved to be more suitable are aluminum-clad, steel-core wires (designated as AW), which are shown in Figure 27-6.

Usually a symmetric umbrella-type top-loading structure is used. However, non-symmetric top-loading could also be used in case of unavailability of ground space around the monopole. It is important to achieve sufficient top-load capacitance, C_0 , to achieve sufficient equivalent monopole height, h_e .

Antenna Insulators

An insulator typically comprises an insulating core that extends between two electrodes that are maintained at significantly different electrical potentials. Insulators have been made with

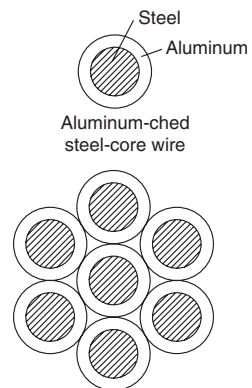


FIGURE 27-7 Cross section geometry of wire conductor

various materials. High-voltage electrical insulators for use at voltages in excess of about 1 kV, and typically at 24 kV, can be made from porcelain, glass fiber, or solid polymeric material. Porcelain has good electrical properties but is heavy and brittle, thus having low resistance to mechanical forces. Glass fiber rod impregnated with an epoxy resin is of lighter weight but generally has inferior electrical performance to porcelain in outdoor applications, or must be environmentally sealed. A composite insulator contains a structurally strong core or rod typically made of fiberglass, a series of electrically insulating sheds, and two metal end fittings crimped onto the exposed ends of the electrical insulator. Electrical polymer insulators are used in power transmission and distribution systems to provide mechanical support for the conductors and to provide electrical insulation between the high-voltage conductors and grounded tower structures. Polymer insulators are made by covering the outer periphery of a core member made of an fiberglass reinforced plastic (FRP) with shed portions made of silicone rubber. Solid polymeric material combines the advantages of light weight and good electrical performance, but has poorer load-bearing properties. Polymeric tubing and/or sheds (annular extensions) may be added to porcelain or glass fiber rods or incorporated into a polymeric insulator for enhanced performance. High-voltage suspension insulators are used to suspend power transmission lines from overhead supports on poles and towers. Suspension insulators used to be made of strings of porcelain insulators having a size and shape required of that material to provide the necessary mechanical strength, dielectric strength, and creepage distance. Nowadays, most suspension insulators are formed of a fiberglass-reinforced polymer rod and an external protective housing forming the weather sheds. The weather-shed housing is usually made of an elastomer or an epoxy material. Elastomer or epoxy weather-shed housings are designed to protect the fiberglass-reinforced rods from weather and electrical activity.

Insulators made of porcelain usually are used in separating active and passive wire elements. Also, cylinder-shaped porcelain insulators are used to achieve base feed in VLF-LF antennas. Other kinds of insulators also have been tried, such as fiberglass-reinforced plastic and oil-filled tubular porcelain also reinforced with a fiberglass strip providing potentially less weight. Three types of insulators are available:

- **Strain type** The porcelain body is subjected to tensile loads.
- **Fail-safe or compression type** The porcelain is supported in a frame that places the porcelain under compression.
- **Stick type** The working loads are taken by a high-tensile fiberglass band located within the oil-filled interior of the tubular porcelain body.

These insulators are illustrated in Figure 27-8.

RF Loss Mechanisms

To achieve high radiation efficiency, the antenna losses should be reduced. Assuming the antenna base current is I_0 , the total consumed power representing the nonradiated part of electromagnetic power is

$$P_{\text{loss}} = \frac{1}{2} R_l I_0^2 \quad (27-12)$$

The losses occurring in the low-frequency antenna and the corresponding contribution to loss resistance are accounted for on the following contributions:

$$R_l = R_{l-w} + R_g + R_{l-c} \quad (27-13)$$

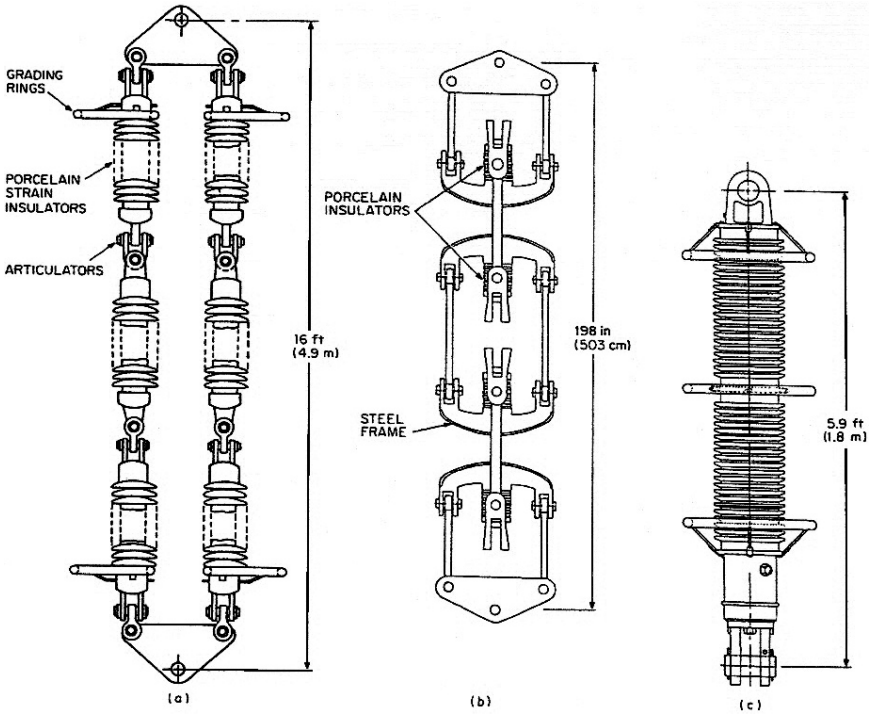


FIGURE 27-8 Three type of insulators: (a) strain, (b) fail-safe, and (c) stick

where

R_{t-w} = Losses related to the or tower and wires (top-loading + supporting wires)

R_g = ground losses due to the imposed fields on the ground surface

R_{t-c} = tuning circuit and transmission line losses at the feeding structure

The conductivity losses of the tuning circuits and mast top-load conductors are computed easily by using circuit concepts and are usually small compared with ground losses.

Assuming the current distribution on the antenna conductors is known, the R_{t-w} , R_{t-c} can be computed using the conductor resistance per unit length ρ :⁷

$$\rho = \frac{1}{a} \sqrt{\frac{f \mu_0}{4\pi \sigma_c}}$$

where α is the conductor radius, f is the radiation frequency in Hz, $\mu_0 = 4\pi \times 10^{-7} \text{ (H/m)}$, and σ_c is the wire conductivity in Siemens/meter (S/m). By computing the total resistance along the wires constituting the antenna structure, the loss resistances can be computed as follows:

$$R_{t-w} \text{ or } R_{t-c} = \frac{\rho}{I_0^2} \int_{\text{conductor axis}} ds I^2(s) \quad (27-14)$$

where the integral is computed along the wire axis.

Computation of Earth Losses

Two mechanisms are involved in losses occurring in the ground medium in the vicinity of the antenna, namely the heating of the ground by horizontal current flow and the capacitive heating of the ground medium because of the vertical electric field imposed on the earth surface. The former loss mechanism prevails, and to compute the ground losses, the horizontal magnetic field near the antenna is computed by using the equation³

$$H_{\varphi} \approx \frac{I_0}{2\pi} \frac{h}{\rho(\rho^2 + h^2)^{1/2}} \quad (27-15)$$

where ρ is the horizontal distance from the antenna feedpoint. The surface loss in case of a surface resistance R_H is computed by a surface integral of the consumed power density on the ground surface:

$$P_H = \iint_{\text{Earth Surface}} d\rho d\phi \cdot \rho R_H H_{\varphi}^2 \quad (27-16)$$

where φ is the azimuth angle. The surface resistance (Ω) depends on the bulk earth conductivity σ (S/m), radiation frequency, f , as well as whether a conductor mesh ground system has been placed around the antenna. Assuming a number of N -conductor electrode system has been placed radially around the antenna following the work of Wait,⁸ the surface resistance can be computed as follows:

$$R_{H, \text{Electr.}} = 10^{-6} \rho^2 N^{-2} f^{3/2} \sigma^{1/2} \quad (27-17)$$

while at the earth medium outside the electrode region it is obtained by

$$R_{H, \text{earth}} = 2 \times 10^{-3} f^{1/2} \sigma^{-1/2} \quad (27-18)$$

Assuming the grounding electrode system has an outer radius of ρ_0 and introducing Eqs. 27-16 and 27-17 into Eq. 27-15, the earth loss resistance is found to be

$$R_g = P_H / I_0^2 = 10^{-6} N^{-2} f^{3/2} \sigma^{1/2} \rho_0^2 / (4\pi) + 10^{-3} f^{1/2} \sigma^{-1/2} (h / \rho_0)^2 / (2\pi)$$

This result is highly valuable because it provides the main loss component of LF antenna systems.

MF Antennas

The ability to achieve antenna sizes having electrical length up to $5/8 \lambda$ improves the antenna radiation efficiency in case of middle-frequency vertical monopole antennas operating at the frequency range 170–1000 MHz. This also allows the decrease of the interference phenomena occurring because of the almost-equal ground and ionosphere waves at 150–300 km at night. Anti-fading antennas attempt to improve the ground wave component by using collinear antennas such as those shown in Figure 27-9. Loading the antenna with a coil at its mid-height provides good anti-fading properties. Depending on the frequency and antenna height, two co-phased half-wavelength or three-quarter-wavelength current distributions could be used. Because of the rather large electrical length of MF antenna towers, top-loading is rarely used.

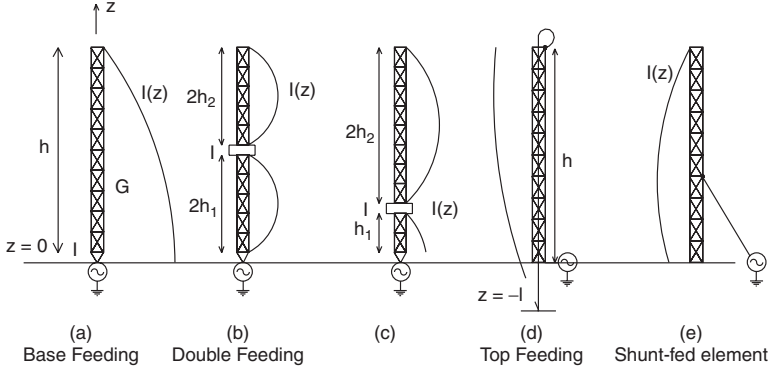


FIGURE 27-9 Various MF antenna design principles

In all practicality, the assumptions of sinusoidal current distribution along the antenna and of the earth's surface to be a perfect conductor provide a very good approximation in computing the radiated far field and the input resistance. In the simplest case of single width feeding assuming the antenna with current amplitude $I(t) = I_0 \cos(\omega t)$, the vertical polarized far field electric strength above the ground surface is

$$E_\theta = \frac{60F(\psi)}{r} \quad (27-20)$$

where

$$F(\psi) = I_0 \frac{\cos(G \sin \psi) - \cos(G)}{\sin(G) \cos(\psi)}$$

is the radiation pattern factor, $G = 2\pi h / \lambda$ is the vertical monopole electrical length, and ψ is the elevation angle as shown in Figure 27-9.

The assumption of sinusoidal current distribution provides the antenna impedance with good accuracy^{7,8} of the real part of the impedance being equal to

$$R_0 = \frac{R_\Sigma}{2 \sin^2(kh)}$$

where the dipole (antenna with $2h$ length) loop resistance is

$$R_\Sigma = 30 \left\{ \begin{aligned} &2[C + \ln(2k_0h) - Ci(2k_0h)] + \\ &\cos(2k_0h)[C + \ln(k_0h) + Ci(4k_0h) - 2Ci(2k_0h)] \\ &+ \sin(2k_0h)[Si(4k_0h) - 2Si(2k_0h)] \end{aligned} \right\} \quad (27-21)$$

where the involved functions are defined as

$$Si(x) = \int_0^x du \frac{\sin u}{u}$$

$$Ci(x) = -\int_x^\infty du \frac{\cos u}{u}$$

and the Euler constant $C = 0.5772$. The reactance of the antenna can also be computed in a similar way.

Another approach to computing the input impedance has been the analysis of the antenna as a lossy homogenous long line with a complex propagation constant $\gamma = \beta + jk$, where $k = 2\pi/\lambda$ is the free space propagation constant and β is the attenuation constant along the antenna.⁹ Then the following relations are valid:

$$Z_1 = Z_0 \left(1 - j \frac{\beta}{k} \right)$$

$$Z_0 = 60(\ln(h/\alpha_e) - 1) \quad (27-22)$$

$$\beta = \frac{R_\Sigma}{h(1 - \sin(2kh)/(2kh))} \quad (27-23)$$

where α_e is the equivalent mast antenna radius. If the mast is being made of an n -polygon multiwire cage, the value of α_e is computed by the equation

$$\alpha_e = a(na_0/a)^{1/n} \quad (27-24)$$

where the terms a_0 and a are defined in Figure 27-10.

The antenna input impedance is computed by the relation of an open-circuited lossy line:

$$Z_{\text{input}} = Z_0(1 - j\beta/k) \coth(\gamma h) \quad (27-25)$$

The analysis of other types of MF antennas shown in Figure 27-9 can be easily carried out based on the perfect conductor ground impedance image theory and assumed sinusoidal current distribution. In Table 27-1 the far field strengths are given as a function of the elevation angle.

Directive broadcasting of MF antennas in broadcasting amplitude-modulated audio signals has been one of the first applications of antenna array theory. The possible directive MF arrays usually are realized by using two or more vertical monopole antennas while the possibility of using horizontal dipole element arrays also has been applied on a limited scale.

MF Vertical Array Antennas

Directional arrays composed of two or more masts are used to provide increased field strength or to minimize interference with coverage regions of nearby MF stations. Mutual coupling between array elements should be taken into account. For a two-element array, gain increase is 2–3 dB, and for four elements it is 4–6 dB. In the frequency region of 525–1605 kHz and power levels 1–2000 kW, many MF stations use array technologies mostly of two elements.

The analysis of MF arrays can be carried out by using the mutual impedance theory of linear antennas.^{7,8} The assumption of perfect conductor ground surface is reasonable since ground wire enrichment is used to achieve high

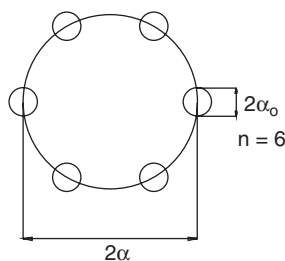


FIGURE 27-10 Definition of the geometric parameters of an antenna mast section

TABLE 27-1 MF Antenna Directivity Characteristics

Antenna Type	Figure 27-9	$F(\psi)$
Base fed	(a) $G = kh$	$F(\psi) = \frac{\cos(G \sin \psi) - \cos(G)}{\sin(G) \cos(\psi)}$
Segmented with two dipoles	(b) $G_1 = kh_1$ $G_2 = kh_2$	$F(\psi) = 2 \left(I_{01} \frac{\cos(G_1 \sin(\psi)) - \cos(G_1)}{\cos(\psi) \sin(G_1)} \cos(2G_1) + I_{02} \frac{\cos(G_2 \sin(\psi)) - \cos(G_2)}{\cos(\psi) \sin(G_2)} \cos(2G_1 + G_2) \right)$
Segmented with a dipole and monopole	(c) $G_1 = kh_1$ $G_2 = kh_2$	$F(\psi) = \left(I_{01} \frac{\cos(G_1 \sin(\psi)) - \cos(G_1)}{\cos(\psi) \sin(G_1)} + 2I_{02} \frac{\cos(G_2 \sin(\psi)) - \cos(G_2)}{\cos(\psi) \sin(G_2)} \cos(k(G_1 + G_2) \sin(\psi)) \right)$
Top fed	(d) $I(z) = I_0 \frac{\sin(k(z+l))}{\sin(k(l+h))}$	$F(\psi) = 2I_0 \left(\frac{(1 - \sin(\psi)) \cos(G_1 + G_2(1 + \sin(\psi))) + (1 - \sin(\psi)) \cos(G_1 + G_2(1 + \sin(\psi)) - 2 \cos(G_1))}{\sin(G_1 + G_2) \cos(\psi)} \right)$
Shunt fed	(e)	Same

efficiency and to prevent raising the elevation of the radiation lobe of the broadcasting antenna. Mostly two-array element systems are used as shown in Figure 27-11. In this case both single-active element or two-active element arrays have been utilized. For two active monopoles the horizontal distance between the two masts is approximately a quarter wavelength, and the two monopoles should be excited by base currents of equal amplitude, with the reflector monopole having a phase lead of 90° . To achieve this condition, the mutual coupling between the two active elements should be taken into account.

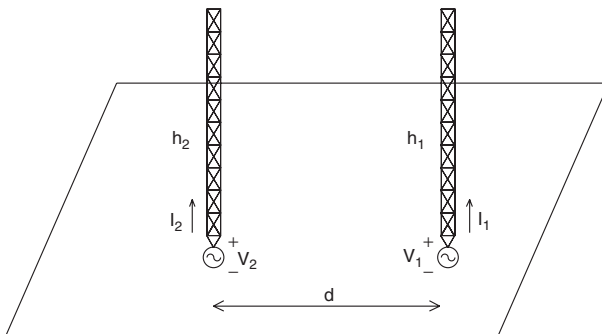


FIGURE 27-11 Two-element directive MF antenna array geometry

The single-active element two arrays are simple, so they usually are the preferred solution in trying to achieve directive beam broadcasting. Assume the setup shown in Figure 27-11 where the monopole element 1 is active. The passive antenna is assumed to have a non-dissipating load with impedance value of jX_{2t} ; then, based on the linear antenna theory, the following equations are written:

$$\begin{aligned} V_1 &= (Z_{11} I_1 + Z_{12} I_2)/2 \\ 0 &= ((Z_{22} + jX_{2t})I_2 + Z_{12}I_1)/2 \end{aligned} \quad (27-26)$$

where V_1 is the driving voltage of the active element, I_1 and I_2 are the currents of the active and passive elements, respectively, while the z -parameters Z_{11} , Z_{12} , Z_{22} , are computed by using standard antenna theory.^{7,8} The input impedance of the active element is obtained easily:

$$Z_{in1} = Z_{11} - \frac{Z_{12}}{Z_{22} + jX_{2t}} \quad (27-27)$$

Furthermore, the ratio of the two currents is

$$m = I_2/I_1 = -\frac{Z_{12}}{Z_{22} + jX_{2t}} \quad (27-28)$$

If the passive element is short-circuited, then $|m| < 1$ always, and as the distance d (see Figure 27-11) increases, the current of the passive element diminishes, since the modulus of Z_{12} decreases. When the passive element is tuned, $Z_{22} = Im(Z_{22}) = -X_{2t}$, and the current in it reaches its maximum value at any inter-element distance d . If d is very small, then the passive element could attain larger current than the active element. For resonance the m ratio value is

$$|m| = |Z_{12}|/R_{22}$$

$$\text{phase}(m) = \pi + \tan^{-1}(X_{12}/R_{12})$$

By varying the distance d and selecting a slightly different height of elements, the passive monopole can operate as a director or reflector. To achieve reflector-type operation, the distance between d should be between 0.2λ and 0.25λ , and the passive element should be slightly higher than the active element. In the past, scaling model measurements have been utilized extensively in designing such arrays. Although scaling model measurements still are used to select and verify design parameters, the availability of highly accurate numerical codes based on method of moments provides more accurate results, as described briefly in Section 27.4.

MF Horizontal Arrays

MF signals can be propagated to long distances (2000–4000 km) by ionospheric reflection, usually during hours of darkness when ionosphere absorption is reduced. For long-distance MF broadcasting, it is possible to use horizontal polarization and also to develop directional radiation patterns by using horizontal dipole array elements. This is useful for secondary coverage areas of MF stations. Usually, primary areas are covered

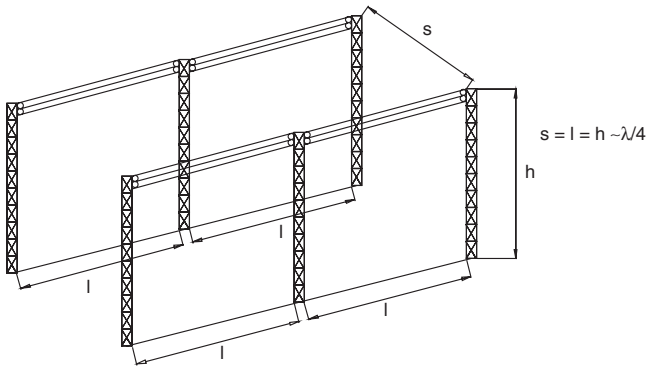


FIGURE 27-12 Horizontal MF array geometry utilizing sky wave transmission

by vertical polarized antennas, which, however, have shortcomings in covering distant regions (>1000 km) because the poor radiation pattern at higher take-off angles limits the signal strength received. In Figure 27-12 the realization of a horizontal array is shown where two broadside arrays parallel to earth surface dipoles are used. No surface-ground wave is induced, and only sky mode is emitted. The take-off angle is a function of the height above the ground of the dipole array at a given frequency.

Antenna Tuning

In all VLF and LF antennas, the antenna height being much smaller than the wavelength of the antenna input presents an impedance $Z_{in} = R + jX$ with few ohms of the real part and a capacitive reactance up to a few thousand ohms. Taking into account the modern solid-state power amplifiers of $50\text{-}\Omega$ output impedance, you need to use matching circuits. In Figure 27-13 a common antenna impedance matching circuit is shown. In all cases a mechanical switch should be placed at the antenna input to protect maintenance personnel during service operations with the antenna system to prevent possible high voltages induced on the tall antenna structure.

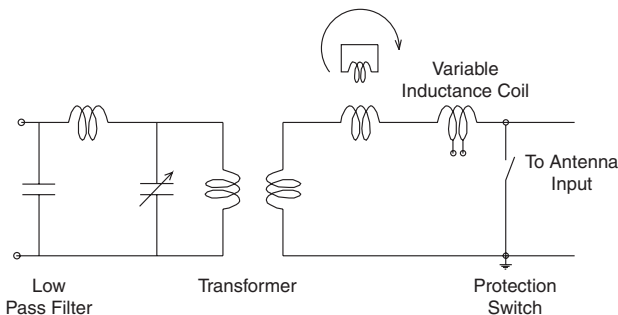


FIGURE 27-13 VLF/LF antenna input impedance matching circuit

As shown in Figure 27-13, a fixed and variable inductive serial network element is used to compensate the highly capacitive load of the antenna. A step-up transformer using low-loss core material is used to increase the low resistance impedance to a higher value near the commonly used 50- Ω transmission line wave impedance connecting the power amplifier with the antenna system. Considering that the distance of the antenna to a 10–100-kW amplifier could be up to 500 m, a matching of better than –20-dB return loss is very important. This requires a coarse- and fine-tuning variation of the matching circuit inductance. A common practice to achieve fine variation of the serial inductance element is to use an air core variable transformer, known as *variometer*, with a short-circuited secondary element. By rotating the secondary element and varying the coupling coefficient of the transformer, a fine change of the inductance is obtained. An additional tuner circuit, such as shown in Figure 27-13, is used between the transmission line and the transformer unit to achieve proper impedance matching of the whole system. To achieve an overall low loss, the antenna tuning circuit elements should be designed and constructed of large-diameter hollow copper wires or even pipes.

An important issue is the protection of the antenna and especially of the power transformer–power amplifier unit from direct lightning strikes on the antenna mast. Gas discharge spark gap protection could be used in regions with high thunderstorm activity. The Tee or Pi-type LC matching networks usually are used in tuning circuits for VLF/LF and MF antennas. In case of power exceeding 50–100 kW, gas-filled capacitors are used to provide protection from induced arcs. All the tuning circuit network elements should be installed into a shielded room that is grounded to the earth and grounding conductor grid.

Environmental Considerations

During the last decades the scientific community and the public have paid increasing attention to possible negative health effects of electromagnetic radiation. Regarding low frequencies and in particular because of the extensive use of industrial electrical energy of 50/60-Hz extremely low frequencies (ELF), there has been extensive study and argument about negative health effects primarily attributed to the magnetic field component of ELF fields. Considering that VLF, LF, and MF fields are quite close to 50/60 Hz, the issue of possible health effects near broadcasting antenna fields are of concern. Various international and government bodies during the last decades have regulated the exposure of the general public and employers to electromagnetic radiation. Presently the most stringent standards are those of European Commission Directive 519-12/7/1999, which sets reference levels concerning the rms electric field strength of 87 V/m and magnetic field strength of 5 A/m. Considering that the vertical monopole top-loaded antenna is the most used antenna and because of the very high capacitive input impedance in the vicinity of the antenna, under the top-loading structure the electric field could have significant values while the magnetic field is comparatively very weak. Therefore, care should be taken to check the electric field near the antenna. Both theoretical and experimental results show that outside of the top-loading region the antenna electric field doesn't violate the 87-V/m standard.

27.4 NUMERICAL COMPUTATIONS OF LOW-FREQUENCY ANTENNAS

The most suitable numerical computation method of analyzing and designing low-frequency antennas is the method of moments (MoM) using a Pocklington integral equation. The analysis of a finite length thin cylindrical dipole antennas has been a fundamental research topic in the history of applied electromagnetism.

In applying the method of moments the radiation problem is formulated by modeling the active parts of the conductor structure by linear overlapping cylindrical segments being smaller than $\lambda/20$. If the number of linear segments is N , then each segment is characterized with its length l_i , ($i = 1, 2, \dots, N$) and radius a_i , ($i = 1, 2, \dots, N$). For mast structures consisting of multisegment tubular cage structures, the radius a_i is computed by Eq. 27-24, while for the top-loading or other wires the actual radii are used. Then the integral equation, taking into account the validity of the condition $a_i \ll \lambda$ and $a_i \ll l_i$, is written as follows:

$$\left(k_0^2 + \frac{\partial^2}{\partial s^2}\right) \sum_{i=1}^N \int_{s=0}^{l_i} ds' K(\vec{r}_j, \vec{r}_i') I_i(s') = -E_{s0}^j(s) \quad (27-29)$$

where the kernel function is the Green's function of the infinite dimension and conductivity of the earth surface averaged over the wire conductor periphery. If the earth medium is assumed to be homogenous and of finite conductivity, the Green's function still is available by taking into account the Sommerfeld integral as an additional term. The s and s' variables represent the *observation* and *source* points on the wire structures, respectively. The right-hand side represents the *imposed* electric fields on the antenna structures and more particularly the applied voltage on the antenna driving points. To determine the unknown current distributions, following the method of moments (MoM) developed during the 1960s–1990s, a set of linear “physical” basis functions are selected to describe the current distribution in terms of a linear sum with unknown weighting coefficients. Then a linear system of equation is obtained after an inner product by a set of “testing” functions of this integral equation of the equation on each segment. In case both “describing” and “testing” belong to the same family, a highly numerically stable procedure is developed, providing accuracy exceeding 1 percent, which is known as the Galerkin method. The selection of physical describing and testing functions require that the engineering inside be pertinent to the problem. An important issue is the stability of the Galerkin method because of the satisfaction of the variational condition, and the accuracy is valid for global quantities such as the radiation pattern and input impedance. Various commercial codes have been developed and are available based on the MoM.

27.5 VERTICAL MAGNETIC LOOP AND HORIZONTAL SLOT ELF AND VLF ANTENNAS USING NATURAL STRUCTURES

Traditional antenna engineering at the low-frequency spectrum mostly uses vertical monopole antennas even as low as 16-kHz radiation frequencies. The limitation imposed by tower heights of 400–500 m makes it rather difficult to construct antennas radiating less than 10 kHz since the achievable efficiencies are very small and the required power levels are excessively high. During the last 50 years various alternatives to vertical monopole antennas have been tried experimentally. Considering all possible small electrical antennas on the perfect conductivity ground surface and derived from electrical and magnetic dipoles and the complementary slot antennas, as illustrated in Figure 27-14, we can state the following radiation modalities:

- Vertical monopole, with its image doubling the antenna electrical length, which has been the basis of low-frequency antenna technology the last 100 years
- Horizontal magnetic loop or electric dipole antennas, both placed near the earth surface and both being non-radiating structures because the image currents are opposite to the antenna currents
- Vertical magnetic loop antenna, which excites an image loop antenna inside the earth, enhancing the primary source
- Horizontal slot antenna, which, as in the preceding case (c), could provide radiating structure

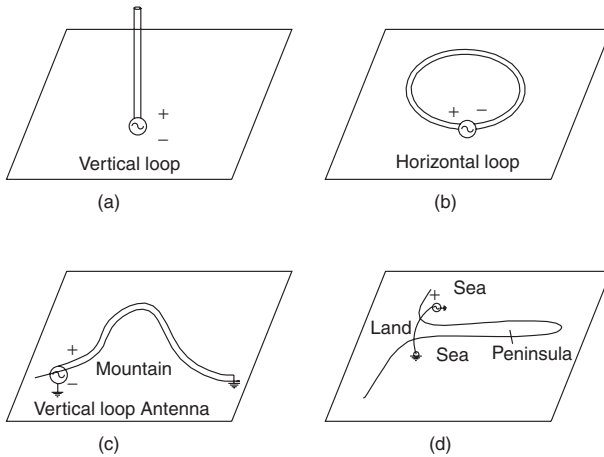


FIGURE 27-14 Possible radiating mechanisms at very long wavelengths

Review of all possible radiating mechanisms apart from the traditional vertical monopole antenna, already analyzed in previous sections, and considering the possibility of building such antennas based on the previously mentioned principles (c) and (d) providing radiation capabilities, gives the following possibilities:

- Vertical magnetic loop, which could be constructed by placing a conductor on a mountain or an island and grounding the ends of the wire. The loop is excited by a serial generator, as shown in Figure 27-14a.
- Horizontal slot antenna, which could be constructed using an island or a peninsula and exiting by a shunt voltage source, as shown in Figure 27-14b.

Both possibilities have been investigated by several researchers during the past 50 years. The best-known case has been the U.S. Navy Extremely Low Frequency (ELF) project, which was proposed in the 1950s by Christofilos, and aimed to emit as low as 70–80 Hz. The vertical magnetic loop was obtained by running parallel wires and grounding them deep inside the earth, which had low conductivity earth in the specific used area. The possibility of utilizing a tunnel and circling a wire around the above mountain has been investigated in New Zealand, and interesting results have been demonstrated in the frequency region of 1–10 kHz providing a proof of concept of a vertical loop antenna. The few reported measurements using vertical magnetic loop antennas confirm the viability of developing broadcasting antennas in the 1–10 kHz frequency region with sufficient efficiency, provided loops with effective area of 1–5 km² are used.

The possibility of developing a horizontal slot antenna using rocky islands and peninsulas also has been investigated by several researchers measuring the input impedance of the wire located in the throat of a narrow island. Measurements of radiated field strength have been carried out on a rocky, narrow (average width 200 m) peninsula of 2.5-km length in the frequency region of 3–10 KHz. Although, based on visual intuition, the possibility of achieving a large size slot radiator is attractive, in practice this is not the case since underground structures penetrated by seawater effectively short circuits the slot radiator and decreases the antenna efficiency to a very small value. Indeed the impossibility of operating the slot mode ELF antenna was predicted in the early 1960s.¹⁰

REFERENCES

1. H. Hertz, *Electric Waves: Being Researches on the Propagation of Electric Action with Finite Velocity Through Space*, Translated by D. E. Jones (London and New York: Macmillan & Co., 1893) and (New York: Dover Publications, 1962).
2. K. A. Norton, *Nature-London*, vol. 135 (1935): 954 and *Proc. Institute of Radio Engineers*, vol. 29 (1941): 623.
3. A. D. Watt, *VLF Radio Engineering* (Oxford: Pergamon Press, 1967).
4. H. R. Bhojwani and L. W. Zelby, "Spiral Top-Loaded Antenna: Characteristics and Design," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (1973): 293.
5. C. E. Smith and E. R. Gnof, "Increased Capacitance for VLF Umbrella Antennas Using Multiple-Wire Rib Construction," *IEEE Trans. Antennas Propagat.* (November 1968): 766–767.
6. H. A. Wheeler, "Fundamental Relations in the Design of a VLF Transmitting Antenna," *IRE Trans. Antennas Propagat.*, vol. AP-6 (1958): 120–122.
7. C. Balanis, *Antenna Theory Analysis and Design*, 2nd Ed., (New York: John Wiley and Sons, 1997).
8. J. R. Wait, "On the Calculation of Transverse Current Loss in Buried Wire Ground Systems," *Apl. Sc. Res.*, vols. B6: 259–275 and B7: 355–360.
9. G. Markov, *Antennas* (English Translation), (Moscow: Progress Publishers, 1965).
10. H. Staras, "Analysis of a Natural VLF Slot Antenna," *Electromagnetic Theory and Antennas*, E. C Jordan (ed.) (Oxford: Pergamon Press, 1963).

BIBLIOGRAPHY

- "Millett G. Morgan, In Memoriam," *Radio Science Bulletin—URSI*, no. 3000 (March 2002).
- Barr R. and W. Ireland, "Low-Frequency Input Impedance of a Very Large Loop Antenna with a Mountain Core," *IEE Proc.-H*, vol. 140, no. 2 (1993).
- Barr, R. "A New Interpretation of the VLF Impedance Measurements of Island Slot Antennas," *Radio Science*, vol. 15, no. 5 (1980): 959–964.
- Burton, R. W., R. W. P. King, and T. T. Wu, "The Loop Antenna with a Cylindrical Core: Theory and Experiment," *IEEE Trans. Antennas Propagat.*, vol. AP-31, no. 2 (March 1983).
- Gould, R. N., "Some Preliminary Experimental Tests of a Novel Method of Radiating at Very Low Frequencies," *Nature*, No. 42773 (April 1961): 332–333.
- Hansen, R. C. "Slot Antenna in a Resistive Screen," *IEEE Trans. Antennas Propagat.*, vol. 46 (July 1998): 1028–1031.
- King, R. W. P., M. Owens, and T. T. Wu, *Lateral Electromagnetic Waves* (New York: Springer-Verlag N.Y. Inc., 1992).
- Kraus J. D., *Antennas*, Chap.13 (New York: McGraw-Hill, 1988).
- Morgan, M. G., "An Island as a Natural Very-Low Frequency Transmitting Antenna," *IRE Tran. PGAP*, vol. AP-8 (1960): 528–530.
- Morgan, M. G., "Comment on a New Interpretation of the VLF Impedance Measurements of Island Slot Antennas by R. Barr," *Radio Science*, vol. 15 (1980): 965–967.
- Uzunoglu, N. and S. Kouridakis, "Radiation of Very Low and Extremely Low Frequencies (VLF & ELF) by a Natural Antenna Based on Island or a Peninsula Structure," *Radio Science Bulletin*, no. 308 (March 2004): 7–12.
- Watt, A. D. "VLF Radio Engineering" (Oxford: Pergamon Press, 1967): 94.

Chapter 28

HF Antennas

Brian S. Collins

*BSC Associates Ltd. and
Queen Mary, University of London*

CONTENTS

28.1	INTRODUCTION	28-2
28.2	SPECIFYING THE RIGHT ANTENNA FOR THE JOB	28-2
28.3	ANTENNAS MOUNTED ABOVE GROUND	28-4
28.4	MAJOR FORMS OF HF COMMUNICATIONS ANTENNA	28-7
28.5	OTHER TYPES OF HF COMMUNICATIONS ANTENNAS	28-12
28.6	HF RECEIVING ANTENNAS	28-14
28.7	BROADBAND DIPOLE CURTAIN ARRAYS	28-16
28.8	SITING HF ANTENNAS	28-20

28.1 INTRODUCTION

The high-frequency (HF) band (2–30 MHz) is widely used for communications and broadcasting. Long-distance communication relies on (virtual) reflection from the ionosphere, a region of ionized particles created by the interaction of the solar wind of high-energy charged particles with upper layers of the atmosphere. The ionosphere is a variable and turbulent medium, and successful communication relies on matching the frequency of transmission and antenna characteristics to its behavior.

There is a substantial difference in practice between public broadcasting operations and point-to-point or point-to-multipoint operations. Broadcast transmissions intended for public reception operate at high power levels to deliver strong signals to distant receivers equipped with simple low-gain omnidirectional antennas. To ensure good audibility, high transmitter powers are used, and these are associated with high-gain directional transmitting antennas whose function is to optimize the signal strength laid down in the target area as well as to conserve spectrum resources by avoiding laying down signals outside the target area. It is common to transmit the same program material simultaneously on more than one frequency band, and the chosen bands may change over the course of a day to match changing propagation conditions. A typical broadcasting station is equipped with a suite of high-gain antennas firing in the directions of interest, often with facilities to modify the beam directions at will in azimuth (and sometimes elevation) direction. Major stations may also be equipped with a rotating high-gain antenna. All the antennas will be of broadband design, allowing them to be used on a number of different broadcast frequency bands, sometimes simultaneously.

International broadcasting is a high-cost activity, with large investments in both transmitters and high-gain antenna arrays and with the high energy costs of operating transmitters with output powers up to 500 kW. The advent of digital transmission permits the reduction of transmitter powers, but this technical possibility will only be secured by international agreement as well as by a change in perception that the station with the strongest signal wins the listeners.

Professional point-to-multipoint services make use of cooperation by the connected stations, and the advent of channel sounders and automatic link establishment (ALE) facilities has led to the availability of reliable communications with relatively low transmitted powers. The advent of digital modulation schemes, adaptive modems, and ALE has provided more reliable communications with increased data throughput, greatly increasing the potential applications of the HF band. *Ionospheric Radio*¹ and the *High-Frequency Radio Automatic Link Establishment (ALE) Application Handbook*² serve as excellent introductions to the subject of modern HF communications.

The amateur radio service has used the HF band for many decades and has proved fertile ground for experimentation with antennas, both of conventional and unorthodox design, especially in the areas of compact antennas and multiband antennas with low to medium gain.

The distinguishing features of HF antennas are their form of construction (usually from wires), their electrical proximity to ground (which determines many of their radiation characteristics), and the way in which antenna performance must be married to a knowledge of the ionospheric path that will support the intended transmission.

28.2 SPECIFYING THE RIGHT ANTENNA FOR THE JOB

The ionosphere has an effective refractive index that depends on height, and an incident electromagnetic wave is progressively refracted. Depending on the ionization density, frequency, and angle of incidence, the signal may pass through the reflecting layer or may be

diffracted through a sufficient angle that it returns to earth. If the incident and emerging ray paths are projected back toward the ionosphere, we can define a “virtual height” that we can regard as being the height of an imperfect reflecting mirror. The ionization density and refractive index are a complex function of time of day, season of year, geographical position, solar activity, and frequency. Four layers are of importance: The *D*-layer (the lowest) absorbs transmissions at MF frequencies during the day, but is transparent when its density falls at night. The *E*, *F*₁, and *F*₂ layers are important for HF transmission and are further discussed below.

The HF band has unique transmission properties and continues to be intensively used. The proper selection of antennas is an important means by which spectrum utilization is maximized. The use of directional transmitting antennas not only increases the signal-to-noise ratio on the wanted path, but it also reduces the signal strength laid down in regions where the signal is not needed and may interfere with other users; directional receiving antennas reduce the extent of interference suffered from other users.

The first step in choosing an antenna is to determine the optimum working frequency (OWF) for the ionospheric path involved. Computer programs are readily available that estimate path losses at different frequencies and the reliability of service for the chosen transmitter power and specified antennas at both ends of the link.^{3,4} The antennas must support the whole range of frequencies that may be needed at any time of day, season of the year, and phase of the 11-year cycle of sunspot activity. The same computer program will identify the probable active ionospheric propagation modes and indicate the corresponding elevation takeoff angles (TOAs) that must be supported by the antennas.

Figure 28-1 shows the path geometry for a one-hop path. The height of the *E* layer is about 100 km. The *F*-layer is more complex in structure and variable in height, particularly with respect to the local solar zenith angle. First-order estimations can be made on the rough assumption that the *F*₁ and *F*₂ layers have virtual heights of approximately 250 km and 400 km. Multiple-hop propagation can be assessed approximately by using Figure 28-1 for each hop. For distances of more than 4000 km, maximum signal results from low-angle transmission in the range from 2 to 15°, with the lower angles generally providing the highest and most stable received power. Propagation prediction programs^{3,4} should be used to establish a link budget and to define the antenna parameters.

Paths of up to about 200 km require high takeoff angles, giving rise to the term *near vertical incidence skywave* (NVIS) paths. The different takeoff angles associated with various path lengths require that we use antennas with different elevation pattern characteristics to serve these paths.

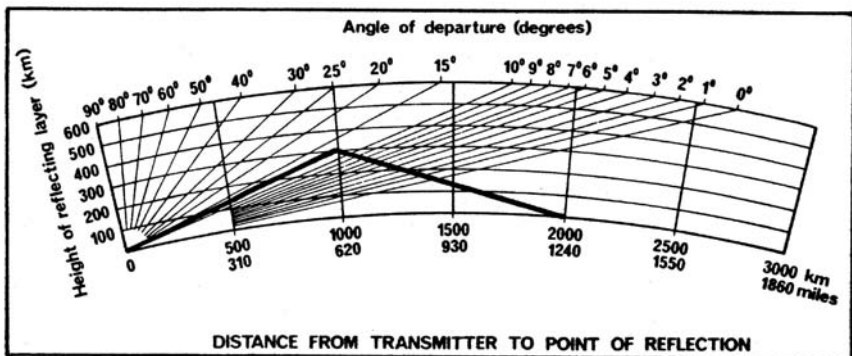


FIGURE 28-1 Takeoff angles for one-hop ionospheric transmission as a function of virtual layer height. The example shows that the TOA for a 2000-km path with a virtual layer height of 300 km is around 12°.

The minimum azimuth beamwidth needed for reliable point-to-point coverage depends on the effects of irregularities and magnetic storms in the ionosphere. Azimuth directions of arrival may typically vary by $\pm 5^\circ$ because of these effects.⁵ This makes the use of antennas with extremely narrow horizontal beamwidths undesirable, especially on circuits that skirt or traverse the auroral zone where much higher angles of off-axis propagation are experienced. For broadcasting or point-to-multipoint services, the beamwidth must cover the target area with allowance for these path-deviation effects.⁶

High antenna gain is required for transmission when large effective radiated power is needed to overcome ionospheric transmission losses. For receiving, high directivity is desirable in the presence of interference, but low radiation efficiencies can often be accepted because in many circumstances the sensitivity of a receiving system is limited by external noise and not by the noise of the receiving system—this is explored later in Section 28.6.

For transmitting antennas, the acceptable input VSWR is usually determined by the characteristics of the transmitter. Modern solid-state transmitters will operate at full output power into a maximum VSWR of around 2:1; above this a transmitter may automatically reduce its output power to limit the voltages and currents in internal circuits.

The power rating of an antenna is determined by both average and peak envelope power (PEP). The average power determines the required current-carrying capacity of the antenna conductors, while the peak power determines the peak electric fields on the antenna, which set requirements on insulator and conductor configurations to avoid arcing or corona discharge, especially in wet weather or at locations where insulators may become contaminated with salt or pollution.

Mechanical requirements are usually determined by wind speed, ice loading, operating temperatures, and corrosive environments. Some applications demand limited antenna size or height, or the whole antenna system may need to be portable for military or emergency relief operations. The tensions in the components of an HF antenna are usually computed using a standard nonlinear finite-element software package. Care must be taken in design to ensure that all joints are fully articulated to avoid failure by the work-hardening of conductors. Insulating synthetic ropes are widely used to support radiating elements and transmission lines, but primary ceramic insulators (typically made from alumina, which has a low dielectric loss factor) are necessary at higher powers to avoid damage by corona discharge.

Many HF antennas are of electrically balanced design and are fed from a coaxial cable through a balun in the form of a transmission line device or a ferrite-cored transformer.

It has become common practice to load HF antennas with resistive or reactive loads; this technique increases the impedance bandwidth but reduces the available gain. The user should always check the power gain of loaded antennas when they are to be used for transmission.

28.3 ANTENNAS MOUNTED ABOVE GROUND

Effect of the Ground

HF antennas are usually operated less than two wavelengths above the ground, whose proximity modifies both the input impedance and radiation patterns. Although accurate evaluation of ground effects is difficult, many properties of HF antennas can be understood by regarding them as being situated over an infinite, flat ground plane. If the ground plane is perfectly conducting, its effect is to create an image of the antenna (see Figure 28-2). For ground of finite conductivity, simple image theory does not give an exact solution, but will often provide useful approximate results.

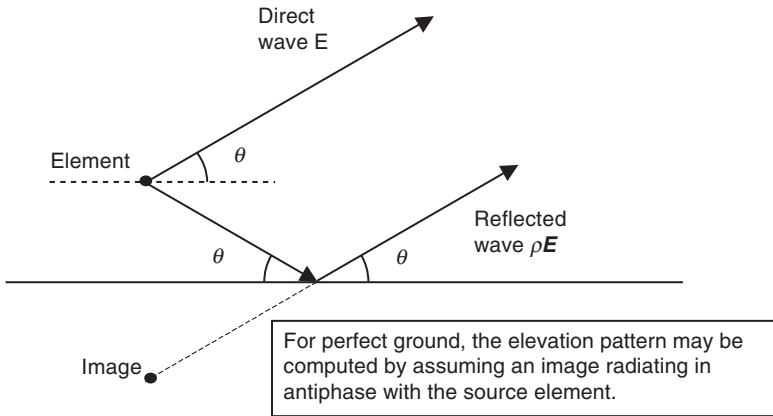


FIGURE 28-2 Direct and ground-reflected rays from an element mounted above ground

Radiation Patterns of Antennas Above Ground

To compute the signal from an antenna at an elevation angle θ° , we must add the ground-reflected wave to the direct wave, allowing for the amplitude and phase of the reflection coefficient of the ground. The proximity of the ground modifies the currents in the elements of an array placed close to it; so although an approximation can be obtained by this method, a full solution for arrays lower than about one wavelength requires the manipulation of the matrix of all the currents in the array including the effect of the ground.

The reflection coefficient for waves with horizontal polarization ρ_h is given by

$$\rho_h = \frac{\sin \theta - \sqrt{(\epsilon_r - jx) - \cos^2 \theta}}{\sin \theta + \sqrt{(\epsilon_r - jx) - \cos^2 \theta}} \quad (28-1)$$

and the complex reflection coefficient for waves with vertical polarization ρ_v is given by

$$\rho_v = \frac{(\epsilon_r - jx) \sin \theta - \sqrt{(\epsilon_r - jx) - \cos^2 \theta}}{(\epsilon_r - jx) \sin \theta + \sqrt{(\epsilon_r - jx) - \cos^2 \theta}} \quad (28-2)$$

where

$$x = 18 \times 10^3 \sigma / f_{\text{MHz}} \quad (28-3)$$

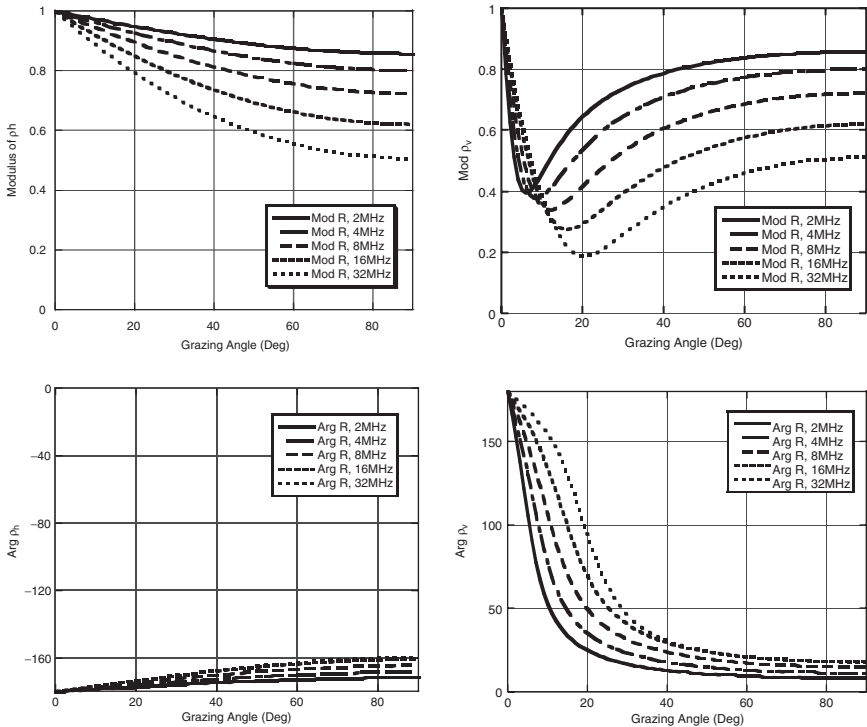
θ = grazing angle

ϵ = relative permittivity of the ground

σ = conductivity of the ground Siemens/meter (S/m)

f_{MHz} = operating frequency in MHz

The ground reflection coefficient for horizontally polarized waves is a well-behaved and slowly varying function. For vertically polarized waves both the magnitude and phase of the reflection coefficient vary with the grazing angle (θ) and the frequency. The angle

Magnitude and phase of reflection, H -polarizationMagnitude and phase of reflection, V -polarization

Reducing conductivity has a similar effect to increasing frequency—see Eq. 28-3.

FIGURE 28-3 Reflection coefficients for horizontal and vertical polarized waves on reflection by ground with $\epsilon_r = 5$ and $\sigma = 10$ mS/m (typical of fertile farmland)

at which the reflection coefficient is minimum is called the *Pseudo-Brewster Angle*, and at this angle the phase of the reflection coefficient is 90° . Figure 28-3 is calculated for $\epsilon_r = 5$ and $\sigma = 10$ mS/m. It will be seen from the form of x in Eq. 28-3 that the effects of falling conductivity and increasing frequency are similar, so over poor ground the values tend to fall even further from complete reflection as the ground absorbs more energy from the incoming wave.

By selecting the polarization and height of the antenna above the ground, we can effectively control the elevation angle of the main beam maximum (the TOA). The ground—even poorly conducting ground—acts as a good reflector at low angles of elevation for horizontally polarized signals, and the gain of a horizontally polarized antenna is effectively increased fourfold when the antenna is located above good ground. (The reader may consider how this comes to be true, as radiation is apparently only confined to one half-space, so why is the gain not merely doubled?) As shown in Figure 28-3, the magnitude of ρ_h is not strongly dependent on ground conductivity at low angles, but the contribution added by reflection from the ground falls off as the TOA increases.

For many types of horizontally polarized antennas the effective height of the radiating region above ground is fixed, with the result that the TOA falls as the frequency increases. This suits the general behavior of HF paths, because long paths with low TOAs generally operate at higher frequencies than high-TOA short paths.

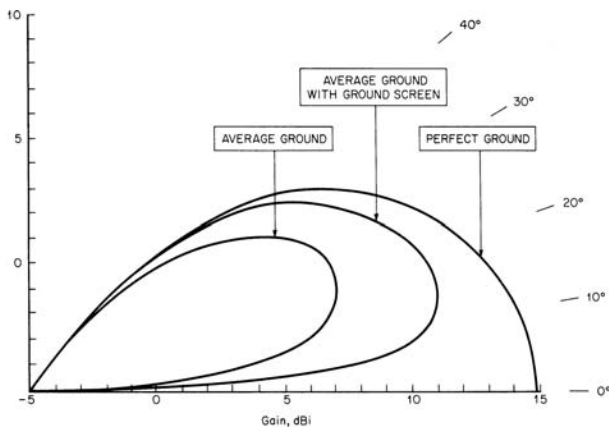


FIGURE 28-4 Elevation patterns of a monopole over ground of different conductivities

The effect on a vertical radiator close to the ground is more complex because the current distribution on the element is modified by coupling to the ground, but the general principle still operates. The finite conductivity of the ground reduces the gain of the antenna at low elevation angles, raising the elevation main beam maximum. This can be seen in Figure 28-4, which also illustrates that the effect of finite ground conductivity can be reduced by the use of a ground mat, in this case comprising wires laid in the direction of propagation and typically spaced by around $\lambda/10$ at the highest operating frequency.

When thinking about the radiation patterns of HF antennas, it is important to remember that the fields in the azimuth plane ($\theta = 0^\circ$) may be very small, and it is usual to specify the “azimuth” pattern for some specified takeoff angle(s)—effectively the cuts of a cone on the solid 3D pattern.

Computer-Aided Design

HF antennas are often designed using computer programs that solve Maxwell’s equations numerically for arbitrary antenna geometries, including the effects of the ground and any ground mat.⁷ They calculate the current distribution on the antenna, and from this they derive the radiation patterns, making allowance for ground effects. In general the radiation patterns are computed to an acceptable degree of accuracy, but the computation of feedpoint impedances should be regarded as a good estimate, rather than reliable fact. It is often worth changing the segmentation of the model and details close to the feedpoint to check that the result is stable and is not sensitive to small changes in the way the real antenna is represented.

28.4 MAJOR FORMS OF HF COMMUNICATIONS ANTENNAS

Many HF antennas comprise derivatives of simple dipoles and vertical monopoles. To meet the requirements of extended bandwidth, these simple prototypes are modified into the forms shown in Figure 28-5.

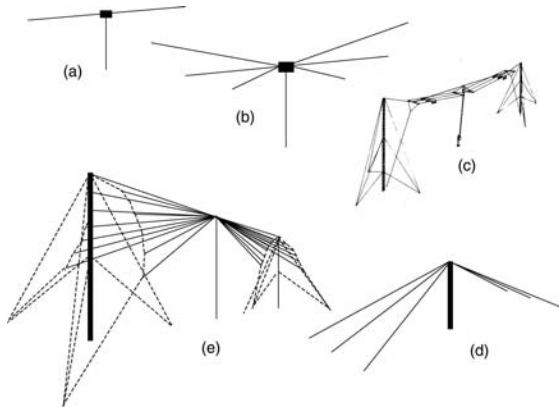


FIGURE 28-5 (a) Dipole; (b) Multiple dipole; (c) Loaded dipole; (d) Fan dipole; (e) 3D biconical dipole

When additional gain is needed, or when it is desirable to form a directional beam, the most common practice is to use some form of log-periodic or curtain array. When an antenna is only required to receive signals, it is possible to use compact low-gain elements and to array these to provide the required directivity to enhance the received signal-to-noise ratio.

Dipoles and Dipole Derivatives

The simplest HF antenna is a wire dipole suspended above the ground (see Figure 28-5a). For NVIS use it should be no more than 0.35λ above the ground to avoid an overhead pattern null. The multiple dipole (Figure 28-5b) provides operation at several separate frequency bands, while the loaded dipole (Figure 28-5c), loaded fan dipole (Figure 28-5d), and biconical dipole (Figure 28-5e) will provide low VSWR over extended bandwidths.

Monopole Derivatives

These designs require a ground mat that forms the second terminal of the antenna and increases effective ground conductivity, enhancing radiation at low takeoff angles. A full ground mat for this purpose comprises a set of between 60 and 180 radial copper (or copper-plated steel) wires, $\lambda/4$ long at the lowest operating frequency, laid on the ground or buried within a few centimeters of the surface (deeper burial leads to losses in the ground above the wires). The inner ends of these wires are joined together at the base of the monopole; they form the second terminal of the antenna and are connected to the outer conductor of the feeding coaxial line.

The biconical monopole (see Figure 28-6) can be constructed using a single metallic mast placed on an insulated base. Internal resonances between the tower and the cone can be removed by connecting the mast to the cone at one or more points. In an alternative form of this antenna the mast is grounded and the lower cone wires are insulated and fed against ground. Midway up the mast the cone wires are connected to the mast to form an inductive loop. This acts as a matching device, allowing the antenna to provide a low VSWR when it is only 0.17λ high. Conical monopoles comprise only the lower half of the biconical antenna, and their mechanical construction can pose some interesting challenges.

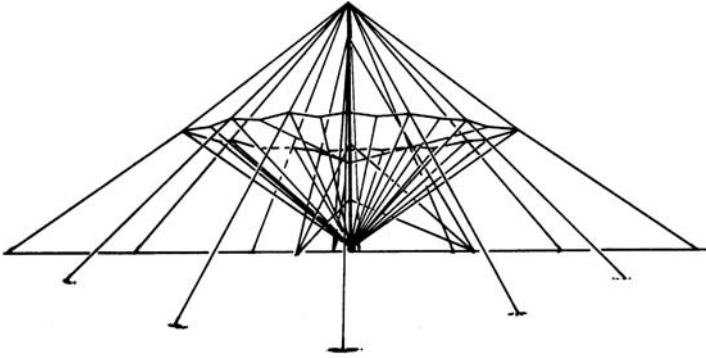


FIGURE 28-6 A biconical monopole antenna

The azimuth pattern of conical and biconical monopoles is circular at all frequencies, but the elevation pattern changes with frequency in the manner shown in Figure 28-4.

Horizontally Polarized Log-Periodic Dipole Array (LPDA)

The basic LPDA (see Figure 28-7) comprises a series of half-wave dipoles spaced along a balanced transmission line and connected to alternate sides. Successive dipole lengths and spacings have a constant ratio τ (typically between 0.8 and 0.95). The bandwidth is limited by the lengths of the longest and shortest dipoles. The number of radiators depends on the frequency range and the value of τ . The parameters τ and α , as discussed in detail in Chapter 13, determine the gain, input impedance, and maximum VSWR of the antenna.^{8,9}

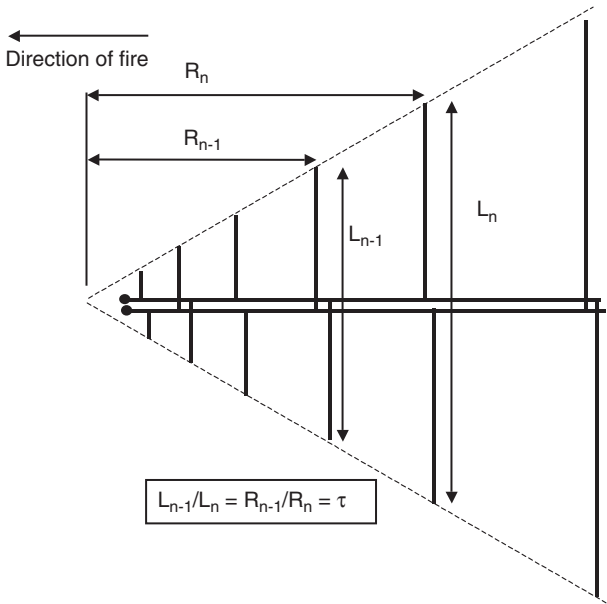


FIGURE 28-7 A transposed-dipole LPDA

At any given frequency only those dipoles approximately $\lambda/2$ long are excited, and these form a single active region. The active region moves along the curtain from the longest to the shortest dipole as the frequency is increased. Because of this movement, it is possible to mount an LPDA so that the electrical height of the active region is constant (the apex of the antenna is placed at ground level) or variable by raising the apex so the active region at the HF end of the array is higher (in terms of the wavelength) than at the lower frequencies. This allows the takeoff angle to be optimized to some extent to match the path characteristics. It is important to choose τ and α carefully because of the possibility of unwanted resonant effects—energy that normally radiates from the active region excites a second active region where the radiators are $3\lambda/2$ long, causing impedance and pattern perturbations.

The antenna is fed close to the shortest element. The balanced feedline is loaded by the elements that modify its effective characteristic impedance. Careful attention to detail is needed if the antenna is to provide a low VSWR across the whole HF band; parameters requiring careful optimization include the characteristic impedance of the feedline and the distances by which it extends beyond the array at both ends. The input impedance of a horizontally polarized LPDA is strongly dependent on the height of the array above ground, and problems can be experienced over sloping ground. (Raising the HF end of the array is usually less problematic than a configuration in which the HF end of the array is electrically closer to the ground.)

A vertically polarized array can be formed by rotating an LPDA into the vertical plane and deforming its shape so the lower end of each element is at the same distance above ground—typically about 1 m. The dipoles are suspended from an insulating catenary attached to a mast positioned behind the longest element; the input end of the feedline is supported on an insulating pole close to the front of the antenna, while the far end of the transmission line is attached by insulators to the supporting mast.

A shorter form of vertically polarized LPDA can be constructed by placing the main feedline parallel with the ground and using only quarter-wavelength radiating elements—effectively a log-periodic monopole array. A practical difficulty with this array is the need to provide the effective 180° phase shift along the feedline between the radiating elements; this has been overcome by various methods including loading or meandering the transmission line between the elements. The complexities of feeding, together with the requirement for a ground mat, have limited the use of this form of array to circumstances in which the height of a standard LPDA is unacceptable—for example, close to airports.

Adequate VSWR performance can be obtained when each radiating element is constructed from a single wire. A long thin element has a high characteristic impedance, and the voltage at the open end is very high even for modest input powers; for peak powers above about 20 kW it may cause corona discharge problems at the open ends. The power rating of the antenna can be increased by forming each element from a number of parallel wires forming a cage perhaps 5–20 cm in diameter. Cage elements have wider impedance bandwidths, so the maximum VSWR of an array can be lower than that possible with single wire elements. Other forms of radiating element and feed systems such as trapezoidal or triangular teeth can be used for the same purposes. A typical horizontal LPDA using half-wave radiators has a gain of 10 to 12 dBi and an azimuth beamwidth of 60° – 80° ; its TOA depends on the height of the active region and can be adjusted by changing the inclination of the antenna relative to the ground. A vertical half-wave LPDA has a typical gain of 11 dBi and an azimuth beamwidth of 110° .

Transformations of the LPDA

The basic topology of the LPDA has proved very adaptable, and the antenna preserves its wideband characteristics even when subject to transformations that provide a variety of

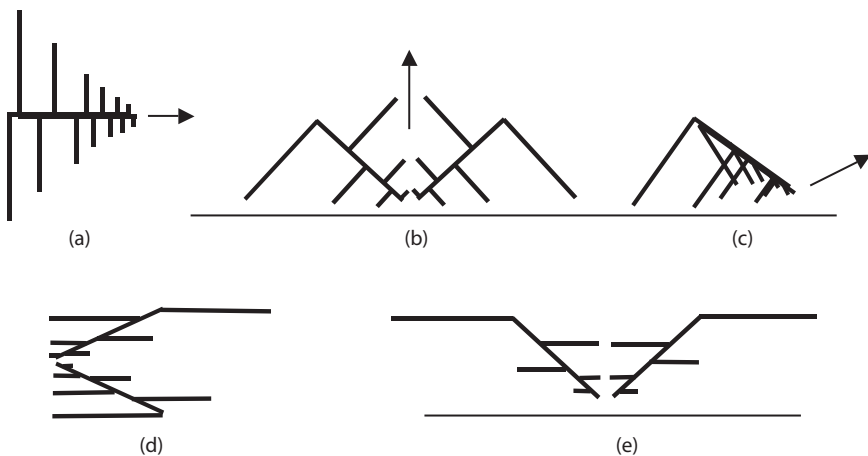


FIGURE 28-8 Transformations of the standard LPDA (a), (b) and (e) NVIS derivatives, using the ground to reflect the downward-directed primary radiation. (c) A common single-mast horizontally polarized version. (d) Sometimes used for high-power rotatable antennas. The arrows indicate the general directions of fire.

new properties. Some of these are shown in Figure 28-8, which shows various ways in which the two half-curtains of the standard LPDA can be mutually rotated and the radiating elements realigned. Horizontally polarized LPDAs can also be stacked above one another to obtain increased gain or arrayed in the azimuth plane to provide omnidirectional or sector coverage.

Rotatable LPDAs

Rotatable LPDAs are often used when the direction of transmission must be changed from time to time and a compact antenna is required—they have been a common sight on the roofs of embassy buildings in capital cities worldwide. For input powers below about 10 kW these antennas often make use of tapered tubular radiating elements but may also be made from wires tensioned by insulated supporting catenary ropes. The antenna has constant physical height above ground, so its electrical height increases with frequency. As a result, elevation grating lobes form at higher frequencies, reducing gain in the principal lobe and introducing nulls at elevation angles that may be required to serve active propagation modes. These antennas are often mechanically complex, with 3D systems of supporting ropes to support the weight of the radiating elements. They are not easy to erect because they are large, heavy, and fragile; they require regular maintenance. The rotating mechanism may be situated at ground level or at the masthead; in both cases careful design is necessary to manage the large torque generated during rotation and periods of high wind.

The large dimensions of a rotatable LPDA can be reduced by loading or folding parts of the longer elements. To minimize antenna dimensions, the highest possible lower-frequency bound should be adopted for any application.

Spiral and Conical Log-Periodic Antennas

A major problem at many HF communications stations is a lack of land available for the required number of transmitting antennas. To help solve this problem, multimode antennas

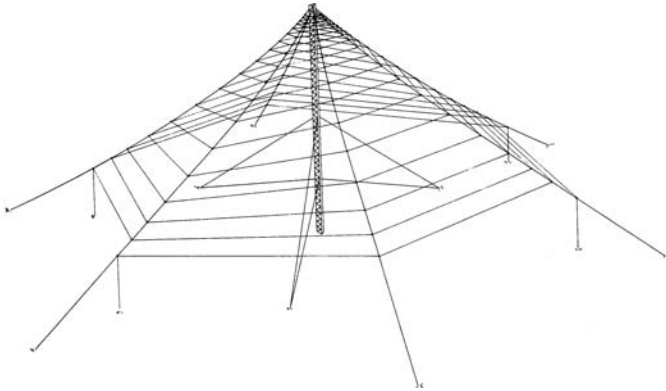


FIGURE 28-9 Multimode spiral antenna (*Courtesy of TCI*)

have been developed that permit several HF transmitters to feed one antenna simultaneously and that provide good isolation between the transmitters. A commonly used structure for such an antenna is a four-arm spiral, as illustrated in Figure 28-9. An antenna of this type operating from 2 to 30 MHz with a gain of approximately 7 dBi can be designed with a height of 37 m and a diameter of about 115 m. Smaller antennas may operate to higher low-frequency limits or may use resistive loading to achieve acceptable VSWR values, reducing their radiation efficiency and power gain.

The four-arm spiral can be excited in four independent modes, three of which are balanced and one that is unbalanced. The balanced modes are horizontally polarized and omnidirectional in azimuth at all elevation angles. The unbalanced mode is vertically polarized and also omnidirectional¹⁰ (see also Chapter 13). In practice the orthogonality of the normal modes can provide interport isolation between 20 dB and 30 dB.

The most common feed system excites the three balanced normal modes. Two of these modes radiate energy directly overhead, and the third mode has a null at the zenith. The configuration shown in Figure 28-9 provides a low TOA at higher frequencies and high TOA at lower frequencies. Excitation of these modes requires that the antenna be equipped with a combination of hybrid and balun transformers.

Some multimode spiral antennas are inverted with their apex located near the ground. This arrangement produces TOAs that are nearly constant with frequency. Other configurations of multimode spirals also excite the fourth (unbalanced) mode, in which the tower and antenna are excited as a vertical radiator.

28.5 OTHER TYPES OF HF COMMUNICATIONS ANTENNAS

Vertical Whip

The vertically polarized whip antenna, usually 2–5 m long, is widely used because it is simple, inexpensive, and easy to mount on a vehicle or the roof of a building. An electrically short whip has a narrow impedance bandwidth and low efficiency; an associated tuning unit is needed to match it at the required operating frequency. The tuning unit may

be automatic in operation, but is not usually fast enough to support frequency hopping, and its input power is usually limited to a few hundred watts. Whips are commonly used on vehicles, but they suffer the disadvantage that their radiation pattern has a deep overhead null, limiting their effectiveness for short NVIS paths. It is common to see whips bent over the top of the carrying vehicle in an effort to reduce the overhead null, but this unfortunately further reduces their bandwidth and efficiency.

Traveling Wave Antennas

A traveling wave antenna comprises some form of leaky transmission line fed at one end and terminated by a resistive load at the remote end. It behaves ideally like a matched transmission line and has a relatively constant input impedance over a very wide bandwidth. Its gain, efficiency, and radiation patterns vary significantly with frequency, reducing its useful operational bandwidth to about an octave.

The rhombic antenna was widely used for point-to-point communication links. It has generally been superseded for transmitting use by the LPDA, which is smaller and requires less expensive supporting structures, and by LPDAs or active loop arrays for receiving.

The sloping-V comprises two wires connected to a feedline at the top of a supporting mast, the wires being arranged in a V-shape in plan with their ends grounded through terminating resistors (each typically 300 Ω). As long as the two legs have an electrical length longer than about two wavelengths, it provides a very simple and quickly deployed antenna with some directional characteristics and modest gain.

The inverted-V comprises a single wire element fed at ground level, suspended above ground at its midpoint, and terminating at ground level in a resistive load. A single or multiple wire ground mat is needed to provide the second terminal for both the input and the terminating resistor. The radiation from a long wire has a main lobe firing obliquely off the axis of the wire; an optimum geometry aligns the wires so the directions of radiation of the two wires coincide. The sloping- and inverted-V, which can be viewed as half-rhombic antennas, find application as quick-erect broadband antennas for transportable operations.

The Beverage antenna comprises a single wire (or more than one connected in parallel) mounted a few meters above ground level. It is terminated with a resistive load at one end and fed at the other end. It provides a simple low-cost directional receiving antenna for use at the lower end of the HF band (or in the MF band).

Loop Antennas

When space is very restricted or an unobtrusive antenna is essential, a small loop can be used as a transmitting antenna. A small vertical loop in free space has a frequency-independent radiation pattern. Its performance can be approximated by a half loop mounted on a conducting plane (see Figure 28-10b). The radiation resistance of a loop is proportional to A^2f^4 , where A is its area and f is the frequency, so the radiation resistance of small loops is very low at low frequencies. The minimum acceptable size is determined by the bandwidth needed at the lowest operating frequency—a bandwidth of 2 kHz can be provided by a loop 1 m high and 2 m wide. The loop has a narrow impedance bandwidth, and it must be tuned at each operating frequency (see Figure 28-10c). A variable capacitor tunes the main radiating loop to resonance while a small coupled drive loop transforms its resistance. It is possible to tune the antenna automatically using a microprocessor-based system that monitors the current in the loops and sets the capacitor to an appropriate value—as with the short whip, a stepper motor is too slow for frequency-hopping applications.

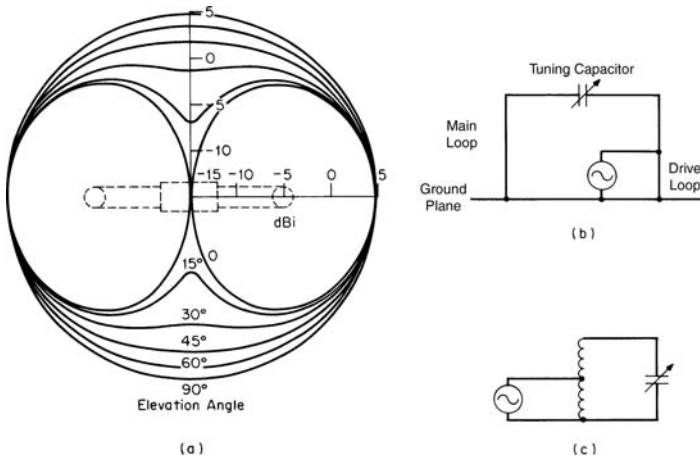


FIGURE 28-10 Vertically polarized transmitting loop: (a) azimuth radiation patterns at different elevation angles; (b) half loop mounted on ground plane; (c) equivalent circuit of half loop on ground plane

The small radiation resistance of the loop at low frequencies makes it inevitable that resistive losses reduce efficiency. Losses are minimized by using a high- Q vacuum capacitor and by ensuring that the current paths in the loop and ground screen have high conductivity. At 3 MHz the efficiency of a well-designed loop will be about 5 percent. Because the radiation resistance of the loop increases with the f^4 but the resistive losses increase only with \sqrt{f} , efficiency rises rapidly with increasing frequency, reaching about 50 percent at 10 MHz and 90 percent above 18 MHz.

There is continuing interest in using loops as NVIS antennas for vehicular use.¹¹

28.6 HF RECEIVING ANTENNAS

The HF environment is generally very noisy, the principal sources of noise being solar and galactic activity, tropical thunderstorms, and human-made noise. If we receive a signal on a very short dipole and gradually increase its length, the resulting signal-to-noise ratio rapidly increases, but once the incoming signal is large enough to dominate internal receiver noise no further increase occurs.

Noise Limitation

F_e is the level of external noise in a 1-Hz bandwidth relative to kT_0 , where $10 \log_{10} kT_0 = -204$ dBW. F_e can be determined from published noise maps¹² or by direct measurement by using a calibrated receiver and antenna. Figure 28-11 shows a typical example of the frequency variation of atmospheric noise; rising noise at low frequencies conveniently compensates for the diminishing effective height of a small loop. Atmospheric noise levels vary with geographical location, season, and time of day. The local human-made noise level depends on whether the receiving site is in an urban, suburban, or rural location; in urban areas it can exceed the atmospheric noise level.

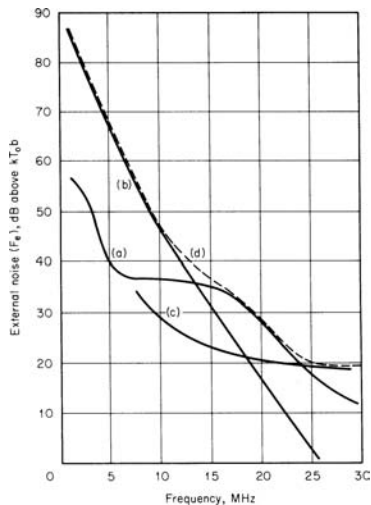


FIGURE 28-11 Example of an external atmospheric and galactic noise figure: (a) atmospheric daytime noise; (b) atmospheric nighttime noise; (c) galactic noise; (d) worst noise condition. For data related to specific locations see “Radio Noise.”¹²

The receiver noise figure F_r is obtained from the receiver specification. The transmission-line noise figure F_l is the reciprocal of its attenuation expressed in dB. For systems with filters, preamplifiers, etc., the noise figure F_{rt} can be determined by computing the noise figure for the equivalent cascaded network.

The antenna noise figure F_a is given by

$$F_a = D - G + L_M \quad (28-4)$$

where

D = directivity, dBi

G = gain, dBi

L_M = mismatch loss = $10 \log_{10} (1 - |\rho|^2)$, dB where ρ is the voltage reflection coefficient

The quantity $D - G$ can be expressed in terms of the efficiency η by using the relation $D - G = -10 \log_{10} \eta$. The efficiency includes resistive losses in both the antenna and the surrounding ground, which may be significant when earth conductivity is low and there is no ground screen.

When $f_e < f_a f_i f_r$, the receiving system is externally noise-limited because the external noise exceeds the internal noise. When $f_e = f_a f_i f_r$, the system is internally noise-limited, and system performance can be improved by reducing the antenna or equipment noise figures. In many geographical locations, atmospheric and noise levels from human activity are extremely low, so receiving systems must have low noise figures to avoid internal noise limitation.

Conventional receiving antennas, designed for transmission, can also be used for reception. These are very efficient, with noise figures often less than 3 dB, making them suitable for use at very quiet receiving sites. Because a high HF receiving-antenna noise figure is often acceptable, a receive-only antenna can be inefficient. Consequently, it can be small and if necessary can be resistively loaded to improve its performance.

Receiving Loops

A practical antenna having a noise figure near the upper limit allowable for most situations is the small vertical loop. A typical loop about 1.5 m in diameter, mounted about 2 m above the ground, has a radiation pattern covering both low and high elevation angles. Successful use has been made of balanced loops (which have classical bi-directional azimuth patterns) and unbalanced loops (which have cardioidal azimuth patterns).

The noise figure of a loop depends strongly on its area and the operating frequency. A loop 1.5 m in diameter has a noise figure of about 50 dB at 2 MHz and 20 dB at 30 MHz. It is common to connect the loop to the coaxial feeder cable using a broadband low-noise amplifier that should have good linearity to ensure that intermodulation products are not produced in the intended signal environment. The amplifier typically has a gain of 10–15 dB and serves to match the loop to the feedline and to establish the receiver system noise figure.

A number of loops can be arrayed to provide a variety of directional radiation patterns. In most arrangements mutual coupling between the loops is negligible, and simple array theory will predict the radiation patterns. Beams are formed by bringing the coaxial cable from each loop into a beamforming network typically constructed using small ferrite-cored hybrid combiners. The most common arrangement comprises eight loops in an end-fire array with spacing between loops of about $\lambda/2$ at the maximum operating frequency. If balanced loops are used, the array can be made bi-directional by splitting the signal from each loop and feeding two beamforming networks. Four bi-directional arrays can be arranged in a circular rosette that provides eight independent beams covering 360° in azimuth. Unbalanced loops have been successfully used to form end-fire, broadside, planar, and circular arrays.

Adaptive Antennas

The availability of cheap microprocessors has made it possible to use active antenna elements as the basis of adaptive antenna arrays. In an adaptive system the amplitudes and phases of the signals from a multi-element array are adjusted dynamically to give a radiation pattern maximizing the wanted signal and minimizing interfering co-channel and adjacent-channel signals. The low mutual impedance between small loops makes them ideal as components of an adaptive array.

Circular Arrays

Arrays of circularly disposed monopoles or LPDAs are used in monitoring or emitter-locating systems (see Chapter 47). A typical array comprises 18 to 36 elements. The output of each element is fed into a beamformer containing delay lines and power splitters forming a narrow azimuth beam. It is possible to generate N equispaced beams simultaneously in an N -element array by splitting the power from each element N ways.

In arrays of LPDAs it is usual to point the elements radially inward. This configuration keeps the effective electrical diameter of the array nearly frequency invariant, making the azimuth beamwidth nearly independent of frequency and eliminating azimuth grating lobes. The disadvantage of the inward-looking array is that each element fires through those opposite, and this complicates the calculation of the radiation pattern.

28.7 BROADBAND DIPOLE CURTAIN ARRAYS

Broadband dipole curtain arrays are used for high-power HF ionospheric broadcasting. A rectangular array of dipoles is mounted about $0.25\lambda_0$ in front of an untuned reflecting screen consisting of closely spaced horizontal wires (see Figure 28-12). This provides an

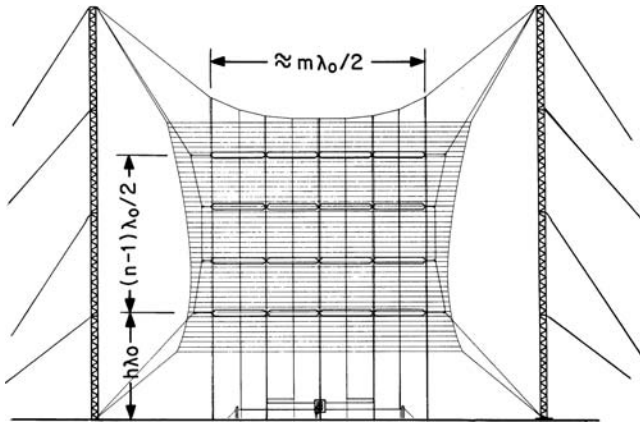


FIGURE 28-12 Basic HF dipole curtain array (antenna shown is designated HRS 2/4/h)

independent choice of azimuth and elevation patterns, the possibility of steering (slewing) the beam direction in azimuth or elevation, high power gain, wide impedance bandwidth, and high power-handling capacity.

Standard Nomenclature

Dipole curtain arrays are described by the internationally agreed nomenclature HRRS $m/n/h$, where

H denotes horizontal polarization

R denotes an array having a reflector curtain

R (if not omitted) denotes that the direction of radiation is reversible

S (if not omitted) denotes that the beam is slewable

m is the width of the horizontal aperture in half wavelengths at the design frequency

n is the number of the dipoles in a vertical stack

h is the height of the lowest dipole above ground, in wavelengths at the design frequency

The design frequency f_0 is approximately equal to $\sqrt{f_1 f_2}$, where f_1 and f_2 are the lower and upper frequency limits; λ_0 is the corresponding design wavelength.

Radiation Pattern and Gain

The TOA (θ_{\max}) and first null elevation angle depend on the average electrical height of the dipoles in the vertical stack and are given approximately by

$$\text{TOA} = \sin^{-1}(f_0 \lambda_0 / 4 f H_{\text{avg}}) \quad (28-5)$$

$$\text{Null} = \sin^{-1}(f_0 \lambda_0 / 2 f H_{\text{avg}}) \quad (28-6)$$

TABLE 28-1 Takeoff Angle of Dipole Array with Reflecting Screen

Number of elements in vertical stack n , half-wavelength spacing	Height above ground of lowest element in wavelengths, h			
	0.25	0.5	0.75	1.0
1	45°	29°	19°	15 and 48 [†]
2	22°	17°	14°	11°
3	15°	12°	10°	9°
4	11°	10°	8°	7°
5	9°	8°	7°	6°
6	7°	7°	6°	5°

[†]Two lobes present

The TOAs at f_0 are given for several configurations in Table 28-1. The lower and upper -3 -dB points are located at approximately $\theta_{\max}/2$ and $3\theta_{\max}/2$, respectively. The level of the minor elevation lobes is determined by the number of dipoles in the stack and their spacing.

The azimuth half-power beamwidth depends primarily on the width of the array, but also depends weakly on the TOA. The beamwidth at f_0 is approximately 76° for arrays one-dipole wide, 50° for two-elements wide, and 24° for four-elements wide, with beamwidths at other frequencies (f_{MHz}) being obtained approximately by multiplying these values by f_0/f . Dipole arrays with adjustable horizontal beamwidth have been constructed that use an RF switching system to excite one or more pairs of vertical stacks.

The gain at f_0 of an array of half-wave dipoles is shown in Table 28-2. The approximate gain at frequencies different from f_0 can be obtained by adding $20 \log (f/f_0)$ to the values in the table.

Slewing

Early forms of dipole curtains used full-wave dipoles. They had narrow-bandwidth feed systems and thin dipoles, so they were capable of operating only in one or two broadcast bands. The radiation patterns were slewed in azimuth by up to 15° using a switched feed arrangement

TABLE 28-2 Gain of Dipole Array with Perfect Reflecting Screen (dBi)

Number of elements in vertical stack n , half-wavelength spacing	Number of half-wave elements wide, m											
	1				2				4			
	Height above ground of lowest element in wavelengths, h											
	0.25	0.50	0.75	1.0	0.25	0.50	0.75	1.0	0.25	0.50	0.75	1.0
1	12.5	13.1	13.9	13.4	13.5	14.3	15.1	14.6	16.1	17.1	17.8	17.4
2	13.9	14.9	15.5	15.7	15.3	16.4	16.9	17.1	18.1	19.2	19.7	19.9
3	15.5	16.3	16.8	17.1	17.0	17.8	18.3	18.6	19.8	20.6	21.1	21.4
4	16.6	17.3	17.8	18.1	18.1	18.8	19.3	19.6	20.9	21.6	22.1	22.4
6	18.2	18.7	19.2	19.5	19.8	20.3	20.7	21.0	22.6	23.1	23.6	23.9

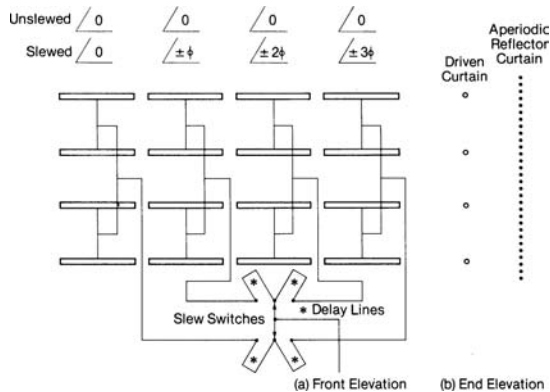


FIGURE 28-13 Slewing system and corporate feed of a broadband HRS 4/4/h dipole array. Dipole length = $0.46 \lambda_0$, dipole spacing = $0.50 \lambda_0$ center to center, and screen-to-dipole spacing = $0.25 \lambda_0$.

to provide a phase difference between the two lateral halves of the array. If greater slew angles are attempted, the horizontal pattern develops large secondary lobes that reduce gain by up to 3 dB and may interfere with co-channel transmissions from other stations. To achieve larger slew angles it is necessary to reduce the horizontal spacing between the dipoles. In modern arrays the dipole length is usually slightly less than $\lambda_0/2$, and the spacing between the stacks is $\lambda_0/2$. The slewing system (see Figure 28-13) enables slew angles to be as large as $\pm 30^\circ$. Systems with up to five slew angles are in common use.

Dipole arrays can be slewed vertically to change their takeoff angle and consequently the distance to the optimum reception area. Vertical slewing is accomplished by altering the excitation phases of the dipoles in each vertical stack using a switching system operating on the elevation feedlines.

Bandwidth

The bandwidth of a dipole array depends on the bandwidth of the dipoles (the fatter the better) and of the feed system. The feed system is usually in the form of a branched tree system (see Figure 28-13) using wideband transmission-line transformers (tapers or multi-section designs).¹³ A well-designed broadband dipole array has a VSWR of 1.5:1 or less in its operating bands. For some arrays and slews, mutual impedances cause some dipoles to exhibit a negative driving-point resistance that can cause large currents to circulate in the feed system. Large circulating currents produce high voltages in parts of the array and rapid changes in antenna input impedance, often accompanied by high input VSWR. Arrays must be designed to eliminate these resonances or to move them to frequencies outside the operating bands.

Practical Considerations

To achieve wide bandwidths, broadband arrays are constructed as fat multiple-wire cage dipoles in the form of rectangular boxes or cylinders. The dipoles are usually folded to raise their input impedances and to provide additional impedance compensation. A folded dipole operates simultaneously in two modes. The radiating mode depends only on the length and

equivalent diameter of the dipole cage. The nonradiating transmission-line mode comprises currents that flow in a loop around the dipole and through the short circuits at the end. The impedance variations of the two modes tend to cancel, so the transmission-line mode can be used to compensate and to tune the radiating mode. This compensation is optimized by moving the folded-dipole short circuit away from the end of the dipole toward the feedpoint.

HRS 4/4 arrays can be designed to have a power rating of 750 kW average (4MW PEP), which is large enough to accommodate two fully modulated 250-kW AM transmitters fed into the antenna simultaneously by using a diplexer. HRS 4/6/h arrays can be designed to handle the power of two diplexed 500-kW transmitters.

The octave bandwidth of a dipole array allows three antennas, 6/7/9/11-MHz, 9/11/13/15/17-MHz, and 13/15/17/19/21/26-MHz, to cover the whole international broadcast spectrum, with multiple-antenna coverage of many of the bands.

28.8 SITING HF ANTENNAS

Sloping and Hilly Terrain

The ground over which an antenna is mounted is seldom flat and uniform. This is of special significance when designing an antenna with a very low takeoff angle. At low elevation angles the zone over which ground reflections add to the direct signal extends far in front of the antenna, and any non-uniformities will affect the shape of the resulting elevation patterns.^{14,15}

Ground Conductivity

For vertically polarized antennas, high ground conductivity is important for the effective operation of transmitting antennas. Vertical transmitting antennas mounted very near seawater, which has a high conductivity (4 S/m), do not require ground screens for radiation-pattern enhancement. For most types of soil, conductivity is significantly smaller than this value, so metallic ground screens are sometimes used when a low takeoff angle is needed. For horizontally polarized antennas, good ground conductivity is not important, although the ground losses below the antenna can be reduced by the use of a screen of wires parallel with the direction of polarization.

Coupling Between Co-Sited Antennas

When energy is coupled between one transmitting antenna and another, it can cause the generation of intermodulation products that can interfere with other spectrum users; it can also cause the incorrect operation of reflected power protection circuits. A typical minimum acceptable coupling is between -20 dB and -30 dB, depending on the technology of the transmitters, but greater isolation will be needed if the transmitter powers involved are substantially different.

Receivers are not tolerant of high signal levels presented at their inputs. When high power transmitters are in use, very high levels of isolation are needed to avoid blocking and cross-modulation being caused in the receiver. The small effective height of active loop receiving antennas results in lower internal intermodulation products than may be provided by an LPDA or other more conventional receiving antenna.

For antennas in each other's far-field region, an estimate can be made using the standard Friis transmission equation:

$$P_r/P_t = G_r G_t (\lambda/4\pi r)^2 \quad (28-7)$$

where

P_r/P_t is the numerical ratio of received to transmitted power
 G_r and G_t are the respective power gains of the two antennas
 r is their separation (m)
 λ is the wavelength

When antennas are cross-polarized in the direction joining them, an additional isolation of 10–20 dB can be assumed. The magnitude of the coupling depends on the polarization of the antennas. For vertically polarized antennas the coupling factor is

$$C_v = 20 \log_{10} (\lambda/8\pi d) + G_t + G_r \text{ dB} \quad (28-8)$$

where

d = spacing between antennas
 G_t = gain of transmitting antenna, dBi
 G_r = gain of receiving antenna, dBi

When antennas are only a few wavelengths apart, the mutually induced currents may cause modification of the expected radiation patterns. If this is likely to be important, a computer analysis may be necessary to assess the possible extent of the interaction.

Side-by-Side Antennas

Horizontally polarized antennas with directive radiation patterns can be placed side by side, often using common towers, with an isolation exceeding 20 dB. Equations 28-7 and 28-8 do not apply in this situation because the antennas are too close. Coupling should be calculated by using computer analysis.

REFERENCES

1. K. Davies, *Ionospheric Radio*, IEEE Electromagnetic Wave Series 31 (London: Peter Peregrinus, Ltd., 1990).
2. *High-Frequency Radio Automatic Link Establishment (ALE) Application Handbook* (Boulder, CO: Institute for Telecommunication Sciences, 1999).
3. "HF Propagation Prediction Method," Rec 533-8, ITU-R, Geneva, 2005.
4. "Computation of Reliability and Compatibility of HF Radio Systems," Rec 842-3, ITU-R, Geneva, 2005.
5. R. Silberstein and F. Dickson, "Great-Circle and Deviated-Path Observations on CW Signals Using a Simple Technique," *IEEE Trans AP*, vol. 13, no. 1 (January 1965): 52–63.
6. A. B. Rogers, C. M. Earrington, and T. B. Jones, "Large HF Bearing Errors for Propagation Paths Tangential to the Auroral Oval," *IEE Proc. on Microwaves, Antennas and Propagat.*, vol. 144, no. 2 (April 1997): 91–96.

7. Numerical Electromagnetic Code (NEC) is a moment-method program developed for the U.S. Navy by Lawrence Livermore Laboratory in 1977. Its derivatives are widely used and can be found on many web sites.
8. E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd Ed. (Englewood Cliffs, NJ: Prentice-Hall, 1968).
9. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design* (New York: John A. Wiley & Sons, 1981).
10. R. G. Corzine and J. A. Mosko, *Four-Arm Spiral Antennas* (Norwood, MA: Artech House, 1990).
11. B. A. Austin and W. C. Liu, "An Optimised Vehicular Loop Antenna for NVIS Applications," 8th Int. Conf. on HF Radio Systems and Techniques, *IEE Conf. Publ. No. 474* (July 2000): 43–47.
12. "Radio Noise," Rec 372–8, ITU-R, Geneva, 2003.
13. G. L. Matthei, L. Young, and E. M. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, Chap. 6 (New York: McGraw-Hill, 1964).
14. W. F. Utlaut, "Siting Criteria for HF Communications Centers," Tech. Note 139, National Bureau of Standards, April 1962.
15. Siting of Radiocommunications Facilities—LF, MF, and HF Transmitting and HF Receiving Facilities, Australian Standard AS 3516.1-1988, Australian Standards Association, Homebush NSW, 1988.

BIBLIOGRAPHY

Further useful general information will be found in:

Handbook on High-Frequency Directional Antennae, CCIR (Geneva: International Telecommunications Union, 1966).

"Transmitting Antennas in HF Broadcasting," BS.80-3 ITU-R, Geneva, 1990.

"Transmitting Antennas for Sound Broadcasting in the Tropical Zone," BS.139-3 ITU-R, Geneva, 1990.

Chapter 29

VHF and UHF Antennas for Communications and Broadcasting

Brian S. Collins

*BSC Associates Ltd. and
Queen Mary, University of London*

CONTENTS

29.1	INTRODUCTION.....	29-2
29.2	DESIGN FOR RELIABILITY	29-2
29.3	SOME UNIVERSAL PRINCIPLES	29-9
29.4	PERFORMANCE, RELIABILITY, COST, AND THE CHOICE OF MATERIALS	29-10
29.5	SAFETY: RADHAZ	29-15
29.6	PROPAGATION.....	29-16
29.7	POINT-TO-POINT LINKS AND BROADCAST RECEIVING ANTENNAS.....	29-17
29.8	OMNIDIRECTIONAL ANTENNAS	29-23
29.9	ANTENNAS FOR BROADCAST SERVICES AND BASE STATIONS	29-28
29.10	BASE STATION ANTENNAS FOR MOBILE PHONE SERVICES	29-36
29.11	SYSTEM CONSIDERATIONS.....	29-41

29.1 INTRODUCTION

The VHF and UHF bands are used for a wide variety of point-to-point and broadcast services with huge economic and social value, generating requirements for many different types of antennas. This chapter includes subjects dealt with in separate chapters in earlier editions of this Handbook and has been extended to cover recent developments in base station antennas, emphasizing methods and techniques that can be applied to these and to many other practical antenna designs. We first review matters of reliability, hardware design, construction, and cost, and then examine the design of antennas for typical applications. Antenna arrays for broadcasting and for mobile communications share many design methods, and engineers working with one of these will find many useful ideas by looking at the methods that have been applied in the other application.

29.2 DESIGN FOR RELIABILITY

Any broadcast or communication system must provide an adequate level of reliability and availability. A link may become unusable if the signal-to-noise ratio falls below the design level; it is important that the design objectives for a system specify the fraction of time for which this may occur. A downtime of 0.01% or even less may be necessary for a link to a lifesaving emergency service, but 1% downtime may be as little as can be economically justified for a radiotelephone in a boat used for leisure-time fishing.

Fading due to statistical fluctuations in the propagation path is usually guarded against by a fade margin in the power budget. In mobile systems it is standard practice to use a diversity system to reduce the impact of fading on system reliability. This takes advantage of the low correlation between fading events over two physically separate paths, at two frequencies, or for two polarizations.

Antenna failure is an important cause of system unavailability. In critical systems reliability is often increased by using two separate antennas fed through a hybrid combiner so that in the event of the failure of one antenna the system continues to operate with a reduced level of signal. When designing an antenna it is very important to understand the user's requirement for reliability and to ensure consistent performance in every delivered unit. These objectives are achieved by understanding and guarding against the mechanisms that lead to failure and inconsistency.

It is very difficult to predict the reliability of an antenna by any formal method, and users will review a proposed design in the light of their experience, looking particularly for weaknesses that may have caused them problems in the past. No telecommunications or broadcasting operator wants antennas to fail in service, and well-designed antennas can provide reliable operation over many years even in difficult climatic conditions. Over-design is expensive and is not necessary; under-design costs more money and endangers vital services. The challenge for engineers new to the field is to develop an intuitive feeling for best practice in this area. The close relationship between mechanical design and electrical function means that antennas must be designed from the outset with their mechanical construction in mind. A good antenna design brings together the best practice in electrical, mechanical, and production engineering.

We now examine important causes of antenna system failure.

Wind-Induced Mechanical Failures

The oscillating loads imposed by wind on antennas and their supporting structures cause countless failures. Aluminum and its alloys are very prone to fatigue failure, and the antenna engineer must be aware of this problem. To achieve real reliability:

1. Examine available wind-speed data for the location where the antenna is to be used.
2. Consider possible local effects such as turbulence around tall buildings or accelerated airflow over steeply sloping ground.
3. Choose suitable materials for critical high-stress components, and use de-rated permissible stress levels to allow for fatigue.
4. Check antenna designs for mechanical resonance.
5. Damp out, stiffen up, or guy parts of the antenna system that are prone to vibration or oscillation.

Base station antennas for mobile radio usually appear to be immune from wind-induced failure, but a large, heavy antenna mounted only on brackets at each end forms a mechanically resonant structure that imposes large fluctuating loads at the support points. The occasional occurrence of bracket failure has led some networks to adopt the practice that every antenna is secured by a stainless steel safety lanyard to prevent a failure in which a roof-mounted antenna could fall into a busy street. More extensive information on design against wind failure will be found in Simiu and Scanlan.¹

Corrosion

The effects of corrosion and wind-induced stresses are synergistic, each making the other worse. They are almost always responsible for the eventual failure of any antenna system. Every antenna engineer should also be a corrosion engineer; it is always rewarding to examine old antennas to see which causes of corrosion could have been avoided by better design. The essence of good corrosion engineering is

- Selection of suitable alloys for outdoor exposure and choice of compatible materials when different metals or alloys are in contact. A contact potential of 0.25 V is the maximum permissible for long life in exposed conditions.
- Specification of suitable protective processes—electroplating, painting, galvanizing, etc.

There is an enormous variety in the severity of the corrosion environment at different locations, ranging from dry, unpolluted rural areas to hot, humid coastal industrial complexes. Detailed information on corrosion mechanisms and control is provided in Uhlig² and Fontana.³

Ultraviolet Degradation

Plastics do not corrode, but they degrade by oxidation and the action of ultraviolet light. These effects are reduced by additives to the bulk materials. The extensive use of plastics in buildings and other outdoor applications has led to the availability of well-stabilized grades of polystyrene (PS), unplasticized polyvinyl chloride (UPVC), acrylonitrile butadiene styrene (ABS), and aminoisoprene-styrene copolymer (ASA). At higher frequencies the lower dielectric loss of PS makes it an obvious choice for radomes, but all these plastics are useful for radomes and other external components.

Fiberglass (GRP—glass reinforced plastic) is widely used in antenna structures and radomes. For radomes it is usual to specify polyester resin and type-E glass to ensure the lowest losses. GRP must always be protected against UV degradation by specifying a suitable stabilizer in the gel-coat. A radome often has a significant effect on the impedance of the antenna elements inside it, so care must be taken to make sure that the potentially high water absorption of polyester resin does not impair the function of the antenna. Vinyl ester resins have much lower water absorption than polyester resins; although slightly more lossy they can be used where water absorption may be a problem.

Water Ingress

Antennas are exposed to the action of rain and of wide daily and seasonal temperature cycling. Water will penetrate incorrectly designed external joints and fill internal voids in antenna or feed components. The designer must ensure that any voids are either self-draining or are hermetically sealed—there is no reliable alternative.

Ice and Snow

The accumulation of ice and snow on an antenna causes an increase in the input voltage standing wave ratio (VSWR) and a reduction in gain. The severity of these effects, caused by the capacitive loading of antenna elements and absorption of radio-frequency (RF) energy, increases as the frequency rises. It is important to check that the antenna and its mounting are strong enough to support the weight of snow and ice that will accumulate on them. This is vital because even when the risk of a short loss of service due to the electrical effects of ice can be accepted, the collapse of the antenna is certainly unacceptable! Ice falling from the upper parts of a structure onto antennas below is a major cause of failure; safeguard against it by fitting lightweight antennas above more solidly constructed ones, or provide vulnerable antennas with shields to deflect falling ice. On tall structures, falling ice creates a danger at ground level; feeder gantries must be robust enough to prevent damage by falling ice, and personnel must be kept clear of areas where falling ice presents a hazard.

In moderate conditions, antennas may be provided with radomes to cover either the terminals of driven elements or whole antennas. In more severe conditions heaters may be fitted inside antenna elements or to prevent the buildup of ice and snow on radomes. A wide range of surface treatments has been used to prevent the adhesion of ice; some of these show initial promise but become degraded and ineffective after a period of exposure to sunlight and surface pollution. Flexible radome membranes and non-rigid antenna elements have been used with some success.

Sand and Dust

Antennas in desert areas are subject to sand-laden wind, which rapidly abrades many surfaces. Fine dust accumulates in any unsealed cavity.

Breakdown under Power

An inadequately designed antenna will fail by the overheating of conductors, dielectric heating, or tracking across insulators. The power rating of coaxial components may be determined from published data, but any newly designed antenna should be tested by a physical power test. An antenna under test should be expected to survive continuous operation at 1.5 times rated mean power and at 2 times rated peak voltage; for critical applications even larger factors of safety should be specified. For multichannel systems with n channels:

$$\begin{aligned} p_{\text{average}} &= p_1 + p_2 + \dots \\ &= np_{\text{channel}} \quad \text{if all channels have equal mean power} \end{aligned}$$

$$V_{\text{peak}} = v_1 + v_2 + \dots$$

$$\begin{aligned} p_{\text{peak}} &= (\sqrt{p_1} + \sqrt{p_2} + \dots)^2 \\ &= n^2 p_{\text{channel}} \quad \text{if all channels have equal power} \end{aligned}$$

In-service failures of high-power antennas often occur at the joints of internal conductors. The passage of high currents through spring connections generates heat, which in turn reduces the temper of the spring, resulting in increased joint resistance and further heating; this leads to a runaway failure. Systems subject to large thermal cycling generate relative physical movement between conductors caused by differential heating and thermal expansion; this movement can cause wear and loss of contact integrity.

External insulating components are subject to the action of water and ice; as well as reducing the surface resistance of the material, water absorption can greatly increase the dielectric loss factor of insulating materials such as fiberglass. Exposed components carrying high surface fields must be designed very conservatively, and materials that are likely to suffer surface tracking should be avoided wherever possible.

Lightning Damage

Antennas mounted on the highest point of a structure are prone to lightning damage. The provision of a solid, low-inductance path for lightning currents in an antenna system reduces the probability of severe damage to the antenna. Electronic equipment is best protected by good antenna design and system grounding, supplemented (for low-power systems) by gas tubes connected across the feeder cables. Figure 29-1 shows a typical system with good grounding to prevent side-flash damage and danger to personnel.⁴⁻⁶

Precipitation and discharge noise are caused when charged raindrops fall onto an antenna or when an antenna is exposed to an intense electric field in thunderstorm conditions. Precipitation noise can be troublesome at the lower end of the VHF band and may be

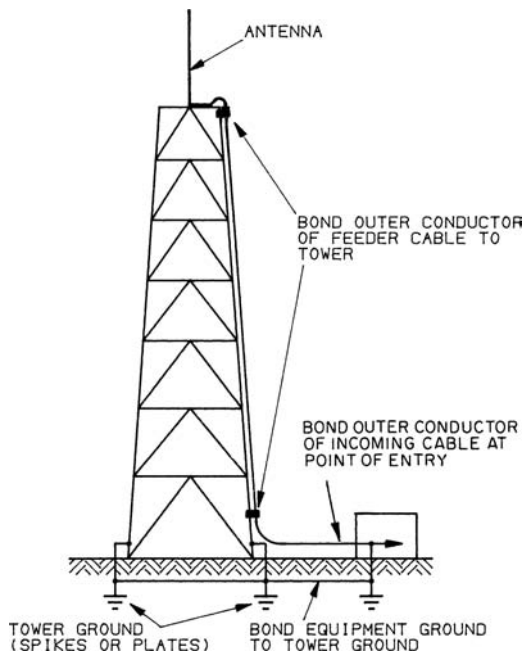


FIGURE 29-1 Typical example of good grounding practice

experienced frequently in some locations. When problems arise, antenna elements may be fitted with insulating covers. These prevent the transfer of charge from individual raindrops into the antenna circuit.

Passive intermodulation products (PIMs) are generated when radio-frequency currents flow across any joint between conductors. They will generally encounter some nonlinearity in the resistance of the joint, caused by the presence of oxide layers on the conductors and made worse by the increase in current density because of the micro-scale roughness of the contacting surfaces. It has been a longstanding practice to use single antennas for the transmission of multiple radio and TV signals, but the only limit on PIM was the radiation of unwanted spurious signals, perhaps 100 dB below the wanted signals. The advent of mobile radio base stations led to the use of a single antenna for transmission of multiple carriers in duplex operation and use of the *same* antenna for multichannel reception. This has led to the usual adoption of the requirement for PIM to be lower than -153 dB with respect to 2×20 W carriers. This ratio ($10^{15.3}$) is approximately the ratio of the distance from the earth to the sun compared with the thickness of a piece of thin paper, so achieving this limit on a consistent production basis presents a major challenge.

The achievement of low PIM depends on following clear rules in design and at all stages of production:

- Avoid any interconductor joint that is not essential to the electrical operation of the antenna.
- Ensure that every essential joint is tightly compressed and that the force on the joint does not depend on any compressible or creep-prone material.
- Avoid dry solder joints—well-soldered joints seem to cause few problems.
- Ensure PCBs are cleanly etched and well washed after processing.
- Protect all essential joints from corrosion (and wherever possible from contact with water).
- Avoid the use of metals with nonlinear conductivity or contact potentials—nickel is a known PIM source, and aluminum joints must be provided with corrosion protection and high contact forces.
- Maintain a high level of cleanliness in storage and assembly areas.

Coaxial Connectors A wide range of coaxial connector interfaces is available, and the choice between them will usually be made on the basis of the mean power ratings that the connectors must support. For permanent installation, most tower-mounted installations use Type N, 7/16-DIN, and the familiar range of bolted EIA flange connectors. If a connector is underrated, the spring inner conductor components will overheat; this leads to softening of the metal from which they are made, reduction of the contact pressure, and consequent increase in contact resistance leading to even more heating and eventual failure. The quality of the fitting of connectors is important in achieving low PIM, low VSWR, and long-term reliability.

For long-term reliability it is standard practice to provide additional weatherproofing to connectors, either by fitting a purpose-designed shrink-on boot or by wrapping with tape. Self-amalgamating tape provides physical protection, but some types do not form a waterproof bond at the connector surface; they should be overwrapped with a layer of petroleum jelly-impregnated tape, which remains waterproof for decades. Where a cable is terminated directly into connection points on the antenna—for example, in domestic radio and television antennas, it is worth applying silicone sealant to the end of the cable and the terminals of the antenna to prevent water penetrating the cable and increasing its loss.

Coaxial Cables For applications at VHF and above, the ratings of cables and connectors are generally determined by their attenuation or the mean power they are required to carry. The maximum power rating of a cable is determined by many factors:

- *The cable diameter* Large cables have lower surface current densities and more surface from which heat can be radiated or conducted.
- *The loss factor of the dielectric* Foam dielectrics have lower losses than the same material in solid form.
- *The thermal conductivity of the dielectric* Heat generated in the inner conductor must escape through the dielectric and the outer part of the cable.
- *The operating frequency* As the frequency rises, the skin depth on both inner and outer conductors diminishes, so resistive losses increase, reducing the ability of the cable to carry power.
- *The ambient temperature* The cable has a fixed maximum internal operating temperature, so in high ambient temperatures it can carry less power.
- *The thermal environment of the cable* Cables installed underground, in floor ducts, or in full sunshine, may require additional derating factors to allow for these conditions. It may be economic to fit a sunshade over a cable rather than to use a larger cable.
- *The maximum operating VSWR* The cable will become hottest at points of maximum current. In general the effect of an operating VSWR σ will be that the maximum power of the cable is reduced by a factor $1/\sigma$. The possibility of operation with a high VSWR caused by fault conditions or ice accumulation must be considered. The VSWR for this calculation should be based on the maximum VSWR into which the transmitter will operate before protective trips reduce its output power. At UHF the axial thermal conductivity of a large cable allows heat to flow away from the points of maximum current, reducing the derating factor.
- *Fault conditions* In fault conditions the power division ratio of a junction may effectively change, increasing the power into parts of the antenna. Any fault that increases the power in part of the antenna must be considered when choosing cable power ratings.
- *Peak voltages* The peak voltage rating of a cable may be exceeded before its mean power rating is reached, so its peak voltage rating or equivalent peak power rating must be checked against the operational requirement. Peak power rating limitations are common at MF and HF and may determine the choice of cables for multichannel VHF and UHF antennas.

Derating factors for these effects are applied cumulatively, so the effective power rating of a cable may be far lower than that given by the power/frequency curve in the cable specification. In a complex antenna system the cables at every level of the feed system must be assessed separately. Information on derating factors will be found in manufacturer's catalogs—often in inconspicuous tables at the back.

Large coaxial cables are heavy and difficult to bend. It is essential to allow adequate bending radii and enough clearance along the cable route to allow it to be installed and formed into position. It is always worth consulting an experienced installer to check the feasibility of complex cable runs. Arrangements for lifting and supporting cables should always be planned by reference to manufacturers' specifications.

Cables for receiving antennas are selected on the basis of their attenuation per unit length (dB/m). The least lossy cables for a given outside diameter are those with helical or foam polyethylene dielectrics. Cables with continuous copper outer conductors have the best screening against pickup when passing through regions of strong

signals—perhaps past transmitting antennas. More flexible cables have braided or tape screens, with multiple screens used for improved shielding when necessary. Some low-cost cables have sparse braids that offer relatively poor immunity to pickup of noise and spurious signals.

Large antenna systems often comprise multiple radiating elements connected by coaxial cables that may be prone to damage by wind vibration, falling ice, and personnel climbing within the array during installation or maintenance operations. To ensure high reliability, secure the cables adequately and route them where they are unlikely to be trodden on or used as “handles.” Staff working on these antennas must be trained to appreciate these issues.

Antenna Design for High Operating Power Many of the factors just listed are relevant when designing an antenna to handle any power that is high enough to cause local heating and damage to its components. A small antenna operating at 2 GHz and carrying 100 W may be vulnerable to power failure if the designer failed to recognize the need to limit power dissipation and to avoid thermally insulating the components that will generate heat under power. A multichannel GSM omnidirectional antenna designed to carry 500 W provides a critical test of the designer’s understanding of thermal design.

For components that will operate at elevated temperatures, make sure the chosen materials will not degrade at their expected operating temperature and that when power is turned on or off, the resulting thermal expansion/contraction can be accommodated without causing stresses. Where thermal expansion may cause relative movement between parts, the design must allow for this by providing a reliable low-resistance current path and avoiding wear by repeated small movements.

Power Testing There is no substitute for conducting a power test, either on a complete antenna or on critical components like radiating elements and power splitters. The mean and peak power levels for the test should be chosen to provide the factors of safety suggested in this section. To provide in-service reliability it is important to understand the operating temperatures that may be experienced by internal components, relating this to the softening point of plastics and to the maximum service temperature for spring connections and solder joints. The tools for temperature measurement include infrared thermal cameras for external components and thermochromic paints and stick-on labels for internal components. Thermochromic labels may heat up by dielectric heating in high RF fields; take care that as far as possible they are protected from them.

When conducting a power test, monitor the temperature of the largest masses to make sure the test is long enough for everything to reach thermal equilibrium. Record the ambient temperature at the time of the test, as the critical parameter is the temperature rise above the ambient temperature recorded at various points of the system; if the antenna will be used in a higher ambient temperature than during the test, allow for this when deciding on service ratings.





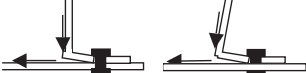









Installation Design and Practice An excellent guide to good practice is provided by the *ETSI Guide*.⁷ Large antennas will be transported in sections and reassembled on site, so all components must withstand transport and rough handling. They must be designed such that reassembly can be carried out on site by an antenna rigger who has no knowledge of the subtle points of antenna design. Components should fit together only one way (the right way) and any (unavoidably) interchangeable parts must be very clearly marked. Any components that were earlier tested as subassemblies should be marked so the same components are reassembled in the same way.

29.3 SOME UNIVERSAL PRINCIPLES

The mechanical design of any antenna system needs to provide a number of basic features:

- Repeatability of electrical characteristics between manufactured samples
- Robustness to withstand transport and installation
- Stability of electrical characteristics in service
- Resistance to lightning damage
- Low RF losses
- Low levels of passive intermodulation products

TABLE 29-1 Typical Mechanical Details Illustrating Good Design Principles for Antenna Design

	Detail	Practical Imperfections	Preferred Detail Example
1	 <p>Strip conductors joined to one another or to ground.</p>	 <p>Contact pressure at circled point is not defined. Loss, power rating, and PIM are uncertain. Impedance changes if contact point moves.</p>	 <p>(a) (b) (b) Can also be achieved by bending or pressing strip end(s).</p>
2	 <p>Right angle connection between flat conductors.</p>	 <p>Any lack of square or deflection of the joint displaces the contact point: loss, PIM, unstable Z.</p>	 <p>Move the fixing closer to the bend, or press the flange to define the contact point.</p>
3	 <p>Connection through a dielectric layer.</p>	<p>Most thermoplastic materials creep under pressure, so the joint integrity is lost and heating occurs. Brittle material may be fractured if joint is over-tightened.</p>	 <p>Add a sleeve over the fastener to remove stress from the dielectric and to form a stable current path.</p>
4	 <p>Clamping through a dielectric layer.</p>	<p>The problems are the same as those of the preceding example, but this is worse because the pressure under the fastener is higher for a given joint closure force.</p>	 <p>Add a sleeve and a washer to reduce pressure: also observe precautions as in (1).</p>
5	 <p>Current path relies on contact through thread.</p>	<p>This is particularly to be avoided. It shares the problems of (3), and the point of current flow between the screw and the upper conductor is undefined. Don't rely on current flowing from a screw into the threaded component.</p>	 <p>Defined current path.</p>
6	 <p>Spring plug contact is smaller than hole.</p>	<p>These examples are exaggerated to show clearly what is happening.</p> <p>Current path is incorrect: impedance uncertain and current density at contact point is high. Plug will overheat.</p>	 <p>Contact is at entry to hole. It is stable and less dependent on spring temper.</p>

Their relative importance varies with the application, but fortunately, the same design methods improve all these desirable features and, with experience, many of them can be incorporated into a design with minimal extra cost, adding real value to the product.

Many important design techniques relate to the configuration of joints between conductors, whether these are part of a radiating structure or an internal feed system. RF currents flow only in the surface of the conductors that carry them, so the achievement of our objectives depends on the way in which the conductor surfaces relate to one another at every interconnection. There must be a well-defined surface current path, maintained independently of production tolerances and bearing sufficient mechanical pressure to ensure a stable current path whatever forces the joint may experience on account of temperature change, vibration, wind load, or other causes. It is essential to remember that in the real world no surfaces are flat—they are always rough and slope in one direction or another. No surfaces are exactly square, and no conductors are completely rigid. Table 29-1 shows examples that illustrate simple principles that must be followed in every antenna design. These examples are not exhaustive. If the reasons for avoiding the details in the left-hand column are appreciated, a whole variety of good solutions may often be found. The criticality of these details increases as the power and/or design frequency increase, or as the requirement for low PIM or resistance to lightning damage becomes more severe.

29.4 PERFORMANCE, RELIABILITY, COST, AND THE CHOICE OF MATERIALS

On many occasions the antenna designer will receive a complete electrical and mechanical specification for an antenna and will create a design that provides the best ratio of performance to cost. On other occasions the designer will have knowledge of the system requirement for the antenna and can optimize the technical and economic aspects of the antenna design together with its supporting structure, cables, and associated hardware; this process often provides more value to the user because the designer can balance a variety of parameters that contribute to the performance and costs of the system.

The reliability of antenna systems continues to increase, reflecting the availability of improved manufacturing techniques, better materials, and the experience of designers and users. There is a huge repertoire of solutions to obtain any needed electrical performance, so much of the antenna designer's time is now spent matching available design concepts to low-cost materials and production methods. Customers' expectations of "more for less" apply to antennas just as they do to any other product. Products that demand hours of patient adjustment to finesse them to meet their specification are not commercially viable; someone will have designed a simple product that works every time—the target we must achieve! In many respects broadcast services operate under the most difficult constraints, as they must provide a very high standard of coverage in a target region while observing stringent restrictions on signals radiated outside the target. This is made difficult because their target regions are often defined by commercial and political constraints that don't fit well with antenna design equations.

The designer must provide the required overall performance for the lowest cost. Reliability of 100% is often very difficult and costly to achieve and is only necessary for a small number of services. By comparison, 99.9% availability will entirely satisfy many users and can be provided much more readily; the user may be unable to justify the high cost of that extra 0.1%.

Cost-effective design is only obtained by:

- Identifying the availability and coverage needed
- Determining the environment at both ends of the path
- Estimating the propagation characteristics of the path and judging the reliability of the estimates
- Selecting equipment and antennas to meet the communications and reliability objectives

This is a general process, applicable to point-to-point and broadcast systems. For a point-to-point system we may have control of the equipment at both ends of the link, and resources can be balanced to provide the best solution. In a broadcast or mobile radio system, coverage plans are based on assumptions about the effective isotropic radiated power (EIRP) of the transmitter, the height and gain of the user's antenna, and the sensitivity of the receiver.

Trade-Offs

For any major scheme it is always advisable to work through the following checklist:

1. Examine the interactions of structure height, transmitter power, feeder attenuation, and antenna gain.
2. Consider using split antennas and duplicate feeders to increase reliability.
3. Consider the use of diversity techniques to achieve target availability instead of a single system with higher power and gains.
4. Review the propagation data, especially the probability of multipath or co-channel interference. Don't engineer a system with 99.9% hardware availability and find 3% outage due to co-channel problems. Check the cost of antennas designed to reduce co-channel problems by nulling out the troublesome signals and the use of diversity techniques to combat multipath fading.
5. Visit the chosen site—if that's not possible, request maps and photographs of the site and its surroundings. General wind data are useless if the tower is near a cliff edge—a careful estimate of actual conditions must be made. A nearby industrial area may mean a corrosive environment, and nearness to main roads indicates a probable high electrical noise level. Look for local physical obstructions in the propagation path. Check access arrangements to confirm assumptions that it will be possible to transport the antenna to site without taking it to pieces.
6. Don't over-design to cover ignorance. Find out!

Cost Issues

A glance through the early chapters of this Handbook demonstrates that many different types of antenna can be used as the basis of design for most applications. The choice between approaches sometimes relates to physical attributes, for example, to minimize dimensions or to ensure robustness. In the majority of cases the designer must compare a number of possible approaches and select candidates for further development on the basis of their suitability for production—the potential for using low-cost techniques and materials and avoiding critical tolerances to ensure reproducibility in production. However hard the path, it is never good enough to produce just “a design that works.”

There are no unique solutions to minimizing costs. The optimum design for any application will depend on labor and material costs in the location where production will take place, although a design that offers the user a higher performance/price ratio may succeed over a simple minimum-cost approach. A design that minimizes the number of components—using each physical component to deliver multiple functions—often provides a result with low cost, high reliability, and clear customer appeal. For a design that is going to be produced in large numbers, figure out the cost of materials, and count the number of parts, holes, fasteners, and critical dimensions. Then rework the design several times to reduce them all!

A major contribution to excess cost is overspecification of mechanical tolerances in an effort to control production variability. There is no substitute for carrying out a sensitivity analysis to check the dependence of a design on component tolerances (both mechanical and electrical)—this allows noncritical parameters to be assigned relaxed tolerances, and the design of critical areas can be reviewed to find less sensitive solutions. It is easy to assume that components manufactured using CNC (computer numerical control) machines have vanishingly small mechanical tolerances, but the designer who relies on this will soon be disappointed. Always ensure that the mechanical engineer who will translate your favorite electrical design into hardware understands which tolerances are critical, which must avoid cumulative errors, and which can be relaxed. Materials that appear interchangeable to a mechanical designer may be very different in electrical properties.

Design for Production Many antenna designs are created to be manufactured in large volumes. Not only must the designer devise a structure with all the right electrical properties—patterns, gain, VSWR, and the rest—but it must be possible to produce the design using standard industrial processes. The performance of a prototype must exceed that required in production by a large enough margin so that almost none of the units produced will be outside the required performance specification. The cost of producing the design in the target volumes, including the costs of scrap and reworking any failures, must meet the requirements of the market and the shareholders.

It is not efficient to begin a design using one technique and collection of materials, and then to decide late in development that to meet production demands the techniques and materials must be changed. An understanding of production techniques that would be economic given the forecast production volumes and rate of production must *precede* electrical design. At every step of electrical optimization the designer must bear in mind the practicability of converting the ideas into producible hardware and must understand the consequences of dimensional tolerances and variable material properties. The design task includes identifying areas of greatest risk and modifying the design to reduce their potential effects; changing materials or processes should only be considered if no cost increase would result.

Materials When selecting materials for antenna construction it is increasingly necessary to consider the need to design for recycling and to ensure that products do not contain toxic materials. The European RoHS directive and initiatives in China, Japan, and some states of the United States have led to the elimination of toxic materials from electrical and electronic products; this trend will continue, embracing a wider variety of products and potential toxins. The approach of many major international companies is to insist on compliance with at least the RoHS limits⁸ for any product they purchase. More extensive checklists follow the Japan Green Procurement⁹ or EIA recommendations.¹⁰ This process is seen as responsible citizenship rather than simply compliance with regulations. The currently restricted materials include lead, mercury, hexavalent chromium, cadmium, PBBs, and PBDEs (compounds previously used as flame-retardants in plastics). As well as prompting the move to lead-free solders, the regulations have changed practices in metal finishing, and

in the stabilization and fireproofing of resins, adhesives, and plastics. Design for recycling is a developing art that will continue to challenge old assumptions about the way materials are selected, components are assembled, and products are packed. Designers everywhere need to be aware of the rules that their companies must conform to when trading in international markets. The Internet is an excellent source of information about current legislation and best practice, both in your country and in international trade.

Conductors The conducting materials used for the construction of most VHF/UHF antennas include galvanized steel, aluminum alloys, and copper (usually in the form of transmission lines and printed circuit laminate). Other materials—including zinc, brass, and stainless steel—are sometimes used in special applications, and also to provide protective coatings. The results of using materials with lower conductivities must be understood in the context of the intended operating frequency and power. At UHF most of the current may flow in the plating rather than the base material, so plating materials must be chosen with their conductivity in mind. Efforts to enhance the conductivity of a less conductive base material are not very reliable; the conductivity of a plated metal (for example, silver) may be lower than that of the material in bulk because of the nature of the crystalline structure of the plated coating.

Insulators Most insulators are made from a surprisingly small range of materials. Antenna system power ratings are often determined by dielectric heating, so for many applications the choice of insulating materials is limited to a small group of low-loss materials: polyethylene (PE), polypropylene (PP), polystyrene (PS, or better, high-impact polystyrene—HIPS), and polytetrafluoroethylene (PTFE). Of these the first three are readily available at low cost and can be molded into complex shapes. PE, PP, and PTFE change shape significantly under prolonged mechanical loading (a process known as *creep*), especially at elevated temperatures. PTFE has a thermal coefficient of expansion that is nonlinear over the usual range of ambient temperatures. PE, PP, and PS suffer surface tracking when exposed to high electric fields—especially if the surface is exposed to rain—and tracking immediately carbonizes the surface, leading to rapid failure. PTFE can be used successfully with significantly higher tangential electric fields along the surface than the other materials. When designing with plastics it is advisable to specify the exact material (including manufacturer and grade). Permittivity and dielectric loss may vary between different grades of the same generic material, especially if recycled material has been included. Ceramic insulators are expensive and are not commonly used in VHF/UHF antenna systems. When using a material for the first time always check its relative permittivity, loss factor, and water absorption.

Printed Circuit Boards These are increasingly used for the construction of antennas at frequencies from the UHF TV band upward. They provide a low-cost method of reproducing large volumes of devices with tight tolerances for use at low and medium powers. The main problems associated with their use are as follows:

Power limitation is caused by lossy dielectric, limitation of current density in narrow conductors, and high field strengths associated with the inevitable sharp edges of etched conducting tracks. When working with laminates based on PTFE or other high-temperature materials, remember that the failure point of the laminate may occur at a very high temperature, so the power rating of an antenna may be limited by the power capable of unsoldering the input connector. Unless provision is made to transfer heat from the critical components, failure may occur at a much lower power than expected.

Performance variability is caused by tolerances in the dielectric constant of the laminate, etching tolerances (over-/under-etch), and problems on long antennas caused by stretching of the photographic film carrying the negative image of the tracks. Errors in

registration between the faces of double-layer structures and tolerances in the positions of holes and other features must not be overlooked. When manufacturing long boards with close tolerances it is important to keep the film negatives for different layers together, so that any stretch does not cause progressive registration errors, and they should preferably be stored flat or laid out flat for some time before they are used. Under- and over-etching cause the width of microstrip lines and other features to vary and can be detected by positioning telltale features at strategic locations. Gap-and-track or wedge features can make very small under-/over-etch clearly visible to the naked eye.

Mechanical stability When designing with PTFE-based laminates it is very important to remember the very poor creep strength of this material, even when reinforced with glass. No electrical connection should rely on compression of the laminate—it will simply flow away, and the integrity of the electrical contact will be lost.

None of these limitations is insuperable, but every designer must understand them, control their extent, and manage their effects. This becomes more critical as the operating frequency is increased. There are few fixed rules, as many of the effects depend on the processes used by the board fabricator and the etching shop. The mobile radio base station market has driven many changes in the cost and variety of available materials and the availability of processors who can cope with boards up to 2.5 m long.

Computer Design Tools

The art and science of antenna design has been hugely advanced by the availability of powerful low-cost computers and specialist electromagnetic simulation packages. For some purposes a computer model will provide a highly accurate estimate of the performance to be expected from a real antenna and allows optimization to meet a practical requirement. Some antennas are still too complex to model as a whole (for example, a large multi-tier TV broadcast array), and the whole design problem must be divided into separate areas for optimization—in this case the design of the panel and the power dividers, the optimization of the geometry and excitation of a tier of the array to create the required azimuth radiation pattern, and the design of the excitation of each tier to provide the required elevation radiation pattern. Simulation is very useful for investigating the effects of mechanical and material tolerances.

As the operating frequency increases, the precision with which an antenna is modeled becomes increasingly important, especially in the input region. Software tools can analyze the performance of an antenna structure but in general are unable to synthesize antennas or arrays to meet specific requirements. A computer does not understand how an antenna works, or which aspects are important to the user; it simply predicts the performance of the input structure.

In the frequency bands addressed by this chapter, moment-method and time domain simulation tools are useful. Full 3D tools are needed for many applications, but 2D or 2½-D tools are useful for some antenna structures and microstrip feed systems.

Computer analysis of aerodynamic and structural design is usually carried out using finite-element methods and allows designers to optimize the shape of critical antenna components. A wide variety of commercial software is available, and every organization engaged in antenna design needs to have access to programs appropriate to the complexity of the work in which it is engaged.

The designer should always view simulation outputs showing the fields, currents, or mechanical stresses present in an antenna design. An intuitive feel for their behavior is one of the hallmarks of the successful designer.

29.5 SAFETY: RADHAZ

The maximum permitted RF field and power flux levels to which personnel may be exposed are specified in various national and international regulations¹¹ and corresponding national standards. These ensure that the effects of the electromagnetic fields on the human body are controlled at a level at which no health effects are known to occur. Those working around LF, MF, and HF antennas should also recognize that the limiting parameter may be the intensity of the H -field rather than E -field or power flux density. Antenna engineers and riggers are probably the most exposed group in the population, and they should be particularly aware of the exposure standards and the need to avoid overexposure of themselves and others.

Antenna engineers become accustomed to working with antennas carrying powers that may be sufficient to cause death or serious injury. Whether climbing close to broadcasting antennas or mobile radio base stations, or conducting power tests in a laboratory, all those working with live antennas need to be made aware of the potential hazards created by non-ionizing radiation. When possible all power should be removed from antennas before climbing near them; transmitters should be locked off, and the person working on the antenna should take the keys or the fuses so that no one can inadvertently apply power while work is in progress. When it is necessary to pass close to operating antennas a competent person should calculate the probable exposure to high fields, and the personnel concerned should carry personal alarms to warn them if they move into areas of unacceptably high field strength. When designing new antennas that are intended to allow internal access (like VHF/UHF panel arrays), care must be taken to make sure that internal access ways are well enough screened to allow safe access, if necessary after reducing normal operating power to permit access. Explicit advisory notices must be placed at the base of the antenna supporting structure to advise workers of the necessary precautions to ensure their safety.

Routine production testing of the passive intermodulation performance of base station antennas creates a significant hazard, and facilities must provide foolproof protection of personnel using the facility and reassurance for those in the vicinity; this can be provided using infrared beams or pressure-sensitive mats to switch off the power in the event that someone is too close to the antenna under test.

Public concerns regarding the possible hazards posed by radio installations appear increasingly likely to be unfounded.¹² Typical power flux densities to which the public is exposed are usually three orders of magnitude below the international limits. Antenna engineers have a responsibility to provide others with accurate estimates of exposure levels and to monitor the developing research literature on potential health effects.¹³⁻¹⁵ Field strengths many wavelengths distant from an antenna—for example, at points on the ground or on nearby rooftops—can be calculated on the basis of the power flux per unit area. The power flux through the surface of a sphere of radius r surrounding an isotropic radiator radiating W_t watts is $W_t/4\pi r^2$ watts per square meter. For an antenna with gain G , the power flux is $GW_t/4\pi r^2$ W/m². By knowing the mean input power to the antenna, its gain in the direction of interest, and the distance to the point of potential exposure, the likely power flux can be calculated and compared with that permitted. At points close to the antenna a simple and reasonably accurate power density estimate is obtained by assuming that W_t watts flow out through a cylindrical surface with the same length as the antenna and a width equal to the distance between the 3-dB points at the relevant distance from the antenna. Assume a back-to-front ratio of only 10 dB for persons close to the antenna, because they cannot be expected to stay in the rear minimum.

29.6 PROPAGATION

The study of propagation effects in the VHF/UHF bands is beyond the scope of this chapter, but all designers of systems required to produce reliable end-to-end connection need to familiarize themselves with tropospheric effects, terrain blocking, and the effects of the built environment. While the prediction of loss over a point-to-point path is relatively straightforward, the prediction of coverage for broadcast and cellular systems has to be approached on a statistical basis to allow for the variations in the position and path conditions to large numbers of users. The possibility of signal impairment by multipath propagation affects many systems. When designing antennas for a long link, consider the possible effects of fading and ducting, in which typical signal levels are either reduced, causing the potential loss of established links, or increased, causing potential co-channel interference from distant users of the same frequency.

Propagation effects influence antenna system design in a variety of ways including the provision of adequate gain and fade margin, the height at which antennas must be mounted, the choice of polarization, and the provision of space, polarization, or frequency diversity. Detailed prediction of coverage areas of broadcast and cellular systems is usually carried out using computer simulations, and a wide variety of planning tools is available, embodying terrain models of various degrees of detail and sophistication. A comprehensive set of graphical data is provided in Shibuya.¹⁶

Use of the finite resource of the electromagnetic spectrum is optimized not only by the adoption of optimum modulation systems and signal formats, but also by the choice of antenna characteristics. Frequency reuse in cellular systems is managed by the positioning of base stations, antenna pattern specifications, and the use of elevation downtilt. The avoidance of co- and adjacent-channel interference in domestic TV systems is managed by the radiation pattern specifications applied to broadcast stations and domestic receiving antennas.

Choice of Polarization

Base stations for mobile services formerly used vertical polarization because it was simple to provide omnidirectional antennas at the mobile terminals and the base station. Cellular systems now often use 45°-slant linear polarization because the signals received from a typical mobile on a pair of antennas polarized at +45° and -45° have sufficiently low correlation to provide useful diversity gain. Circular polarization is often used for radio services to reduce the influence of the polarization of the receiving antenna and to make it easier for the untrained user to find a consistent high-quality signal.

There is sometimes an advantage in using horizontal polarization for obstructed point-to-point links in hilly terrain, but the choice of polarization is often determined by the need to control co-channel interference. Orthogonal polarizations are often chosen for antennas mounted close together in order to increase the isolation between them.

The use of circular polarization (CP) can reduce the effects of destructive interference by reflected multipath signals, so CP should be considered for any path where this problem is expected. CP has been used with success on a number of long grazing-incidence over-sea paths where problems with variable sea-surface reflections had been expected to be troublesome. Both ends of a CP link must use antennas with the same sense of polarization. In some countries circular polarization is used for TV services, but for this application the receiving antenna is usually horizontally polarized.

29.7 POINT-TO-POINT LINKS AND BROADCAST RECEIVING ANTENNAS

Most low/medium gain antennas are of end-fire design. When higher gain is needed, it is usually obtained by forming arrays of basic end-fire antennas. The antenna styles described here are only a small subset of all the possible designs, but they represent a large proportion of antennas used worldwide for practical applications.

In some countries, bowtie dipole and corner reflector designs are popular for domestic TV reception, and they are described later in this section. From a cost/performance point of view they have little significant advantage. Simple log-periodic antennas will cover the UHF band. A variety of hybrid Yagi-Uda, log-periodic, and other solutions is used to provide multiband operation over the VHF and UHF bands. Many of these designs have only modest gain, but the high signal levels available outdoors in many cities allow them to provide adequate reception. In situations with co-channel interference or multipath (ghosting), a more conventional antenna with a narrower azimuth beamwidth and lower sidelobes is often preferred.

Yagi-Uda Antennas

The Yagi-Uda antenna is the most successful general-purpose directional antenna design at frequencies up to at least 2.5 GHz. It is inexpensive and simple to construct, and will provide gains of up to about 17 dBi (or more if a multiple array is used). Its bandwidth can be chosen to suit the application; almost an octave bandwidth can be achieved, but at the expense of a lower gain than can be provided over a narrower band. At low frequencies the realizable gain is limited by the physical size of the antenna; in the upper UHF band a reflector antenna may be simpler, less costly, and more reliable if a large gain is required.

Yagi-Uda antennas provide unidirectional beams with moderately low side and rear lobes. The characteristics of the basic antenna can be modified in a variety of ways, some of which are shown in Figure 29-3. The basic antenna (*a*) can be arrayed in linear or planar arrays (*b*). When the individual antennas are correctly spaced, an array of N antennas will have a power gain N times as large as that of a single antenna, less an allowance for feeder losses. Table 29-2 indicates typical gains and arraying distances for Yagi-Uda antennas of various sizes. Different array spacings may be used when a deep null at a specified bearing is required, but the forward gain may be slightly reduced.

To achieve closely definable radiation patterns it is important to feed the radiating element using a suitable balun to avoid unwanted radiation from the feed cable. Typical Yagi-Uda antennas will provide cross-polar discrimination (XPD) of at least 20 dB in the main beam direction; higher XPD can be obtained by observing good symmetry and a well-balanced feed/balun system.

Yagi-Uda antennas can be built to support high input powers, and they are commonly used for directional broadcast transmission. Only one element and its drive circuit must be specially designed to support high input power.

The bandwidth over which the front-to-back ratio is maintained may be increased by replacing a single reflector rod by two or three parallel rods (*c*)—as often used for domestic TV receiving antennas. The back-to-front ratio of a simple Yagi-Uda antenna may be increased either by the addition of a screen (*d*) or by arraying two antennas with a quarter-wavelength axial displacement, providing a corresponding additional quarter wavelength of feeder cable to the forward antenna (*e*). A well-designed screen will provide a back-to-front ratio of as much as 40 dB, while that available from the quadrature-fed system is typically 26 dB.

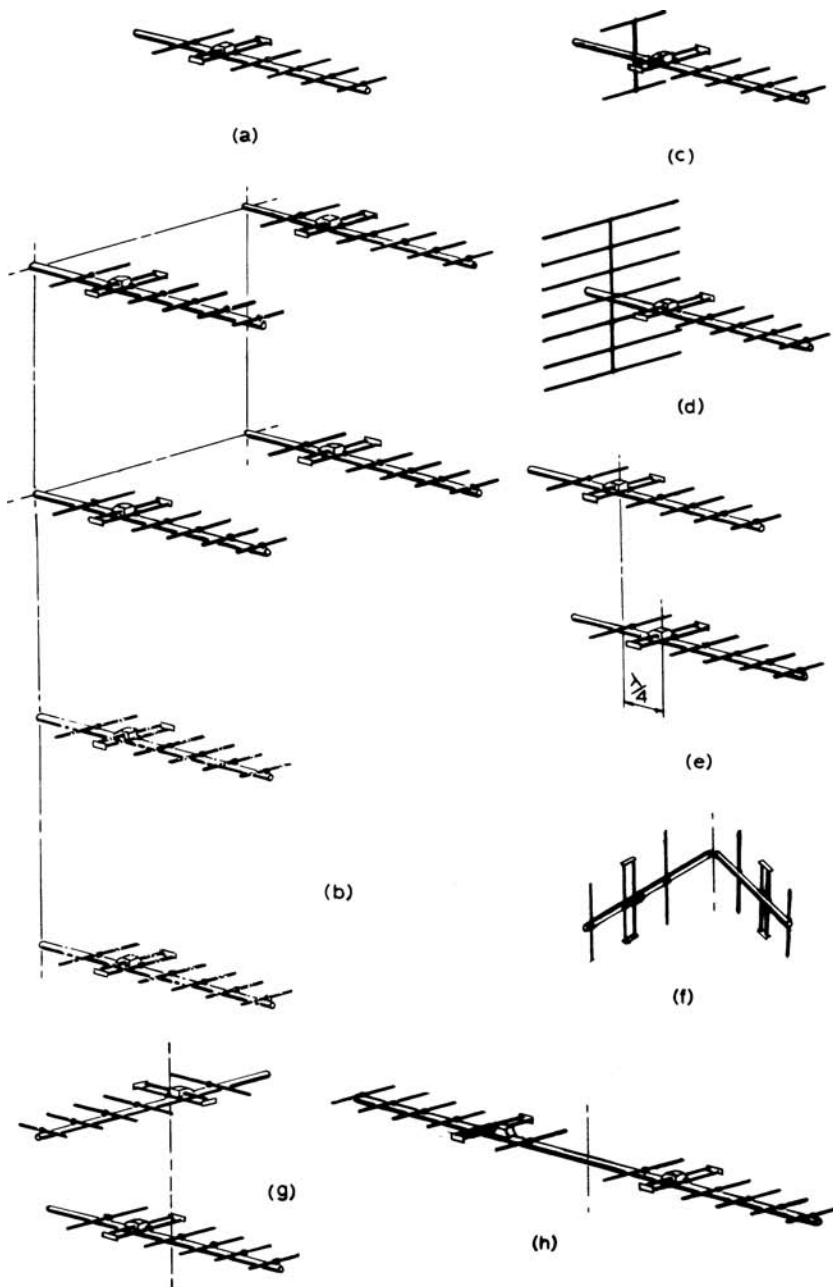


FIGURE 29-2 Configurations of Yagi-Uda antennas: (a) Standard six-element antenna; (b) Stacked and bayed arrays; (c) Double reflector rods; (d) Reflector screen; (e) Increased F/b ratio by $\lambda/4$ offset; (f), (g), (h) Arrangements to produce azimuth radiation patterns for special applications

TABLE 29-2 Typical Data for Yagi-Uda Antennas

Number of Elements	Typical Gain, dBi	Spacing for Arraying, λ
3	7	0.7
4	9	1.0
6	10.5	1.25
8	12.5	1.63
12	14.5	1.8
15	15.5	1.9
18	16.5	2.0

Circular polarization can be obtained by using crossed Yagi-Uda antennas: a pair of antennas mounted on a common boom with their elements set at right angles. The two antennas must be fed in phase quadrature or be fed in phase and mutually displaced by a quarter wavelength along the boom. (Using the in-phase method the equality of the phases and current amplitudes on the two components is less disturbed by the VSWR of the individual antennas as long as they are identical.)

There has been some interest in slow-wave end-fire arrays that use long, closed forms for their elements, such as rings or squares. They have not proved popular in practice, probably because they have narrower bandwidths than an optimal Yagi-Uda antenna and are more complex to manufacture. The domestic TV receiving antenna market is sometimes subject to fashion, and some interesting local variants are seen in cities round the world.

Log-periodic antennas are widely used for applications in which a large frequency bandwidth is needed. The gain of a typical VHF or UHF log-periodic antenna is about 10 dBi, but larger gains can be obtained by arraying two or more antennas. The disadvantage of all log-periodic designs is the large physical size of an antenna with only modest gain. This is because only a small part of the whole structure is active at any given frequency.

The most common design used on the VHF-UHF bands is the log-periodic dipole array (LPDA). (See Chapter 13 and Kraus and Marhefka.¹⁷) After selecting suitable values for the design ratio τ and apex angle α , the designer must decide on the compromises necessary to produce a practical antenna at reasonable cost. The theoretical ideal is for the cross-sectional dimensions of the elements and support booms to be scaled continuously along the array; in practice, the elements are often made in groups by using standard tube sizes, and the support boom is often of uniform cross section. The stray capacitances and inductances associated within the feed region are sometimes troublesome, especially in the UHF band, and can be compensated only by experiment. The coaxial feed cable is usually passed through one of the two support booms to avoid the need for a wideband balun.

Printed-circuit techniques can readily be applied to LPDA design in the UHF band and above, as the antenna is easy to divide into two separate structures that can be etched onto two substrate surfaces. At the lower end of the VHF band the dipole elements may be constructed from flexible wires supported from an insulating catenary cord.

A typical well-designed octave-bandwidth LPDA has a VSWR less than 1.3:1 and a gain of 10 dBi. Log-periodic antennas have been widely used both as TV receiving antennas and also as transmitting antennas at low-power relay stations in the VHF and UHF broadcast bands. LPDAs are produced commercially for surveillance and measurement applications with bandwidths exceeding a decade. Other log-periodic designs that can be found in Chapter 13 are less common in commercial application in the VHF/UHF bands, probably for reasons of cost and windload.

Helices A long helical antenna has an easily predicted performance and is simple to construct and match. A VSWR as low as 1.2:1 can be obtained fairly easily over a frequency bandwidth of 20%, and wider bandwidths are possible if the helix is tapered or stepped in diameter. Conductive spacers may be used to support the helical element from the central support boom, so the antenna can be made very robust. At higher frequencies it may be more convenient to support the helical element by winding a conducting tape onto a dielectric rod or tube. The maximum gain that can be obtained from a single helix is limited by the physical length that can conveniently be supported, typically ranging from 12 dBi at 150 MHz to 20 dBi at 2 GHz.

Helices can be arrayed for increased gain; to obtain correct phasing the inputs must be driven in phase, and the start position of each helix in the array must be the same. Low gain helices—in particular, short quadrifilar helices—find application as reflector feed units and as antennas for small satellite terminals.¹⁸ Ceramic-loaded quadrifilar helices have been used as receiving antennas for GPS systems.¹⁹ For further information on the design and performance of helices see Chapter 12.

Bowtie Dipoles and Reflectors

Arrays comprising a column of bowtie dipoles mounted in front of a reflecting screen are common as horizontally polarized domestic TV receiving antennas in some countries. The triangles can be approximated with wire mesh, provided the mesh spacing is less than one-tenth wavelength, or by a fan of rods connected at the feedpoint. Flare angles between 60° and 80° combined with a half-length up to 0.58λ provide satisfactory input impedance and radiation patterns. To provide more gain and an increased front-to-back ratio, dipoles are usually stacked vertically and mounted approximately 0.25λ in front of a reflecting screen made from parallel rods spaced 0.1λ apart (at f_{\max}). In this configuration the dipoles are fed by an open balanced transmission line. The measured gain for one and two dipoles over a flat screen is also shown in Figure 29-4. Commercially available antennas typically have a VSWR less than 2.0, a front-to-back ratio greater than 15 dB, and sidelobe levels less than 13 dB below the peak gain over 90% of the UHF band.

Corner Reflectors Well-designed corner-reflector antennas are capable of providing high gain and low sidelobe levels, but below 100 MHz they are mechanically cumbersome. Before using a corner reflector, make sure that the same amount of material could not be more effectively used to build a Yagi-Uda antenna, or perhaps a pair of them, to do the job better.

In the UHF band, corner reflectors can be very simply constructed from solid or perforated sheet and a variety of beamwidths and back-to-front ratios obtained by the choice of the apex angle, spacing of the dipole from the vertex, and the width of the reflector. The apex of the corner is sometimes modified to form a trough (see Figure 29-5). The provision of multiple dipoles extends the antenna aperture and increases the available gain.

Panel Antennas An antenna that comprises a reflecting screen with simple radiating elements mounted over it, in a

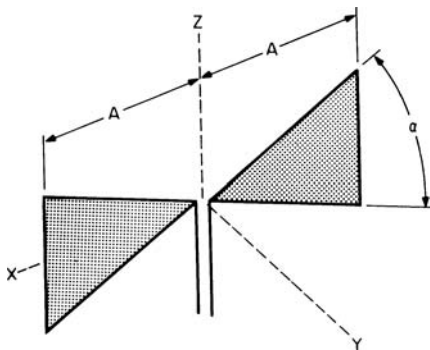


FIGURE 29-3 Bowtie dipole

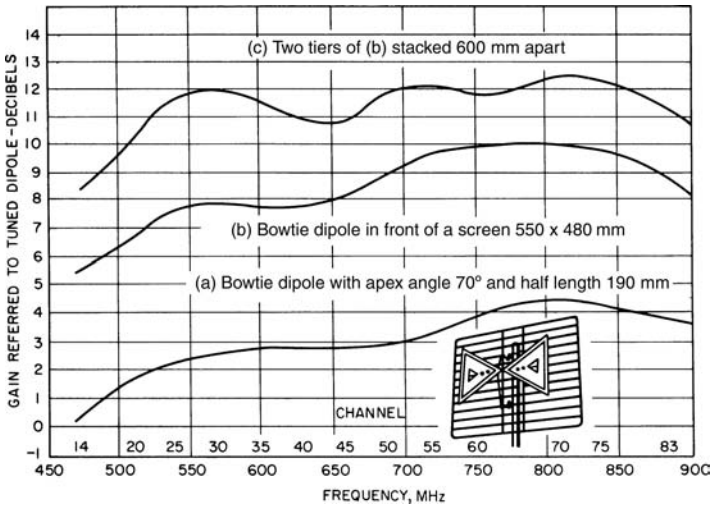


FIGURE 29-4 Gain characteristics: (a) Bowtie dipole ($\alpha = 70^\circ$, $A = 190$ mm); (b) Bowtie dipole in front of screen (550×480 mm); (c) Two tiers stacked 600 mm apart in front of a screen

broadside configuration, is generally termed a *panel antenna*. An array may comprise one or more panels connected together.

Typical panels use full-wavelength dipoles, half-wave dipoles, or slots as radiating elements (see Figure 29-6). For some applications they have advantages over Yagi-Uda antennas:

- More constant gain, radiation patterns, and VSWR over a wide bandwidth—up to an octave
- More compact physical construction—the phase center is maintained closer to the axis of the supporting structure, providing better control of the azimuth radiation pattern
- Very low coupling to the mounting structure
- Low side lobes and rear lobes

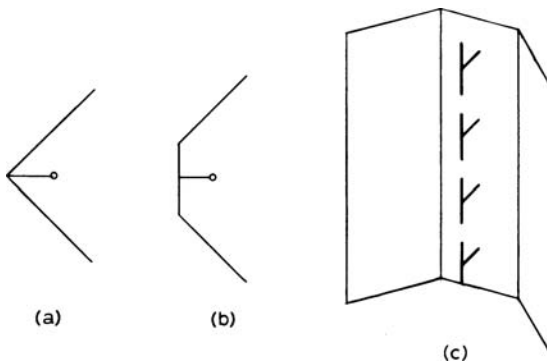


FIGURE 29-5 Corner and trough reflectors

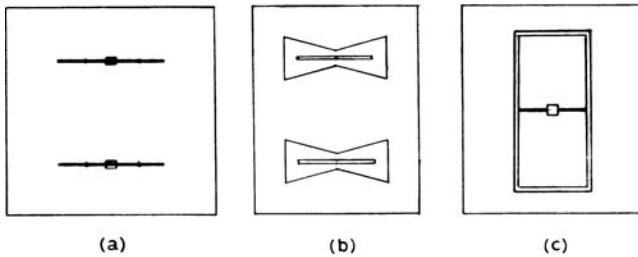


FIGURE 29-6 Panel antennas: (a) Two full-wave dipole elements; (b) Two batwing slot elements; (c) Skeleton-slot elements

Panel antennas for frequencies in the UHF band lend themselves to printed-circuit design methods, as the radiating structures, feedlines, and matching system may all be produced by stripline techniques. At lower frequencies the radiating elements are often mounted at voltage minimum points using conducting supports, so a strong, rigid construction can be produced. A solidly built but lightweight panel for a military application is shown in Figure 29-7a. Here an all-welded aluminum frame and a skeleton-slot radiator are used so that the antenna will resist rough use in the field. Panels are often used as building blocks for complex arrays for radio and TV broadcasting.

Paraboloids The design of a high-gain antenna may be reduced to a problem of illuminating the aperture necessary to develop the specified radiation patterns and gain. The size of the aperture is determined only by the gain required, whatever type of element is used to fill it. As the cost of the feed system and the radiating elements doubles for each extra 3-dB gain, a stage is reached at which it becomes attractive to use a single radiating element illuminating a reflector that occupies the whole of the necessary antenna aperture. The design task is reduced to choosing the size and shape of the reflector and specifying the radiation pattern of the illuminating antenna. If the antenna aperture is incompletely filled or its illumination is non-uniform, the gain that is realized decreases. The ratio of the achieved gain to the gain obtainable from the same aperture when it is uniformly illuminated by

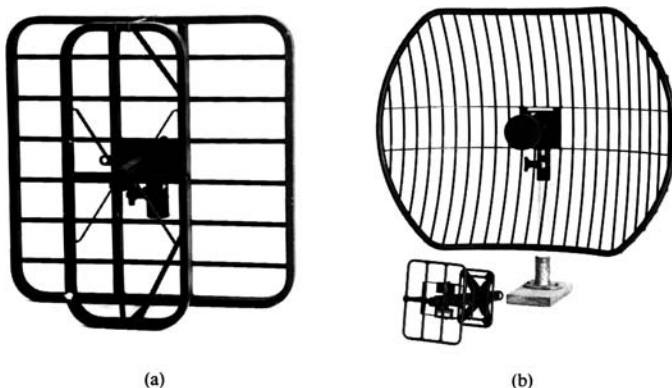


FIGURE 29-7 Robustly constructed antennas for military use: (a) Skeleton-slot-fed panel (225–400 MHz); (b) Grid paraboloid (610–1850 MHz) (Courtesy of Jaybeam Ltd.)

TABLE 29-3 Typical Paraboloid-Antenna Configurations

Frequency (MHz)	Diameter (m)	f/d ratio	Construction	Feed Type
200	10.0	0.5	Mesh paraboloid	4-element Yagi-Uda
700	3.0	0.25	Solid skin	Dipole and reflector
900	7.0	0.4	Perforated steel sheet	Horn
610–960	1.2	0.25	Grid of rods	Slot and reflector
1500	2.4	0.25	Solid skin	Dipole and disk
1350–2500	1.2	0.25	Grid of rods	LPDA

lossless elements is termed the *aperture efficiency* of the antenna. In a receiving context, this quantity represents the proportion of the power incident on the aperture that is delivered to a matched load at the terminals of the antenna.

In the VHF and UHF bands, a reflector may be made of solid sheet, perforated sheet, wire mesh, or a series of parallel curved rods. As the wavelength is large, the mechanical tolerance of the reflector surface is not very demanding, and various methods of approximating the true surface required are possible. Table 29-3 indicates some of the combinations of techniques currently in use and illustrates the diversity of the methods that are successful for various purposes.

Grid paraboloids are attractive to produce because the curvature of all the rods is exactly the same; only their length varies across the antenna. A typical example is shown in Figure 29-7b. The main deficiency of grid paraboloids is the leakage of energy through the surface, restricting the front-to-back ratio that can be achieved. For example, at 1.5 GHz a front-to-back ratio of –30 dB is a typical limit. If a greater front-to-back ratio is needed, it may be possible to adopt an offset geometry. Alternatively, the reflector bars can be extended in depth, or an orthodox continuous skin of solid or perforated sheet can be used in place of the grid; the consequent increase in weight and wind-loaded area must be accepted as a necessary penalty for improved electrical performance. Electrically small reflectors have poor radiation pattern performance because the feed has physical dimensions independent of the reflector diameter, so feed blockage is more troublesome and leads to high sidelobes.

Radomes are frequently fitted to feeds or complete antennas in order to reduce the effects of wind and snow. They may be made from fiberglass or in the form of a tensioned membrane across the front of the antenna. In severe climates it is possible to heat a radome with a set of embedded wires, but this method can be applied only to a plane-polarized antenna.

Point-to-point links using tropospheric-scatter propagation require extremely high antenna gains and typically use a reflector that is an offset part of a full paraboloidal surface constructed from mesh or perforated sheet. Illumination is provided by a horn supported at the focal point by a separate tower. For a full discussion of the design of reflector antennas refer to Chapter 15.

29.8 OMNIDIRECTIONAL ANTENNAS

Simple Low-Gain Antennas The simplest types of antennas will provide truly omnidirectional azimuth coverage only when mounted in a clear position on top of a tower. Figure 29-8 shows standard configurations for ground-plane and coaxial dipole antennas and demonstrates that these forms are closely related. They are cheap and simple to

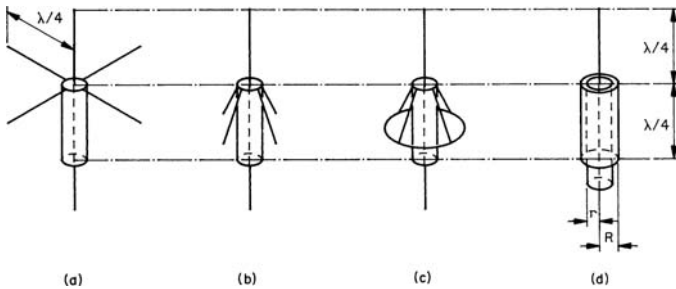


FIGURE 29-8 Low-gain base-station antennas: (a) Standard ground plane with radials; (b) Ground plane with sloping radials; (c) Ground plane with closed ring; (d) Coaxial dipole

construct, and may be made to handle high power. Exact dimensions must be determined by experiment, as the stray inductance and capacitance associated with the feedpoint insulator cannot be neglected. The use of a folded feed system can provide useful mechanical support and gives better control over the antenna impedance. (Both the relative diameter of the feed and grounded conductors and the point of their interconnection can be varied.) The operational bandwidth of the coaxial dipole d depends critically on the characteristic impedance Z_0 of the coaxial section formed by the feedline (radius r) inside the skirt (radius R). If this section has too small a Z_0 , radiating currents will flow on the outside of the feeder line unless the skirt length is exactly $\lambda/4$. The impedance, gain, and radiation pattern of the antenna then become critically dependent on the positioning of the feedline on the tower, severely limiting the useful bandwidth of the antenna.

Discone Antennas The discone and its variants are the most commonly used low-gain wideband antennas. The useful lower frequency limit occurs when the cone is a little less than $\lambda/4$ high, but the upper frequency limit is determined almost entirely by the accuracy with which the conical geometry is maintained near the feedpoint at the apex of the cone.

Discones may be made with either the disk or the cone uppermost. The support for the upper part of the antenna usually takes the form of low-loss dielectric pillars or a thin-walled dielectric cylinder, fitted well outside the critical feed region.

Variants of the basic discone use a bicone in place of the conventional cone and replace the disk by a short cone with a large apex angle. At the lower end of the VHF band discone antennas can be mounted at ground level, so a minimal skeleton disk that couples to the ground may be used if some loss of efficiency and the propagation effects associated with a low antenna elevation can be accepted.

Collinear Arrays The ground plane and coaxial dipole have about the same gain as a half-wavelength dipole. When more gain is needed, the most popular omnidirectional antennas are simple collinear arrays of half-wave dipoles. The original array of this type is the Franklin array shown in Figure 29-9a. This design is not very convenient owing to the phase-reversing stubs that project from the ends of each half-wave radiating section, but various derivatives are widely used. The arrangement at Figure 29-9b uses non-inductive meander lines to provide phase reversal; that at c is a rearrangement of the original, while those at d and e use coaxial line sections. Arrangements such as these may be mounted in fiberglass tubes to provide mechanical support, and the designs at b , c , and f are suitable for production by printed-circuit techniques. In each of these arrangements, the elements are connected in series; an input-matching section transforms the input impedance of the lower section, which may be $\lambda/2$ or $\lambda/4$ long, to 50 Ω . A set of quarter-wavelength radial

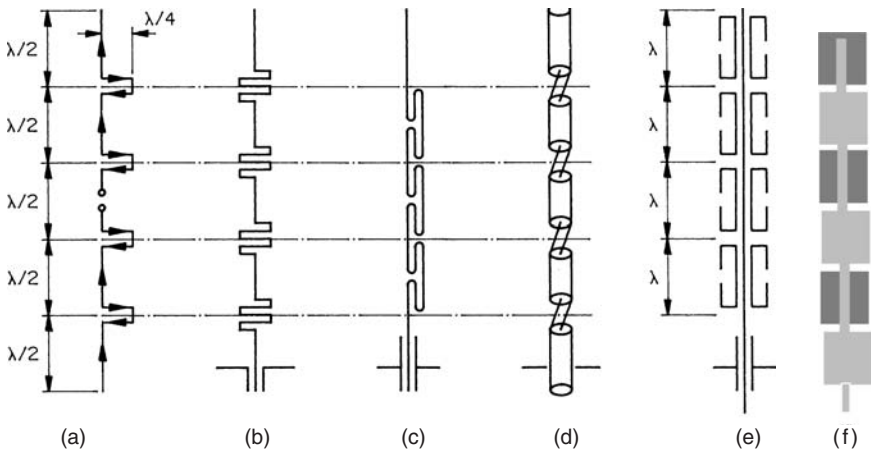


FIGURE 29-9 Collinear dipole arrays: (a) Franklin array; (b) Array with meander-line phase reversal; (c) Array with transposed coaxial sections; (d) and (e) Alternative coaxial forms; (f) Planar version

elements or a quarter-wavelength choke is used to suppress currents on the outside of the feeder cable. The maximum useful gain available from these arrays, typically about 10 dBi, is limited by two factors:

- There is mechanical instability in a very long antenna with a small vertical beamwidth.
- The available excitation current diminishes away from the feed as a result of the power lost by radiation from the array.

In the case of the coaxial-line designs, each section is shorter than a free-space half-wavelength so that the correct phase shift is obtained inside the section. The examples shown would typically provide a gain of 9 dBi at the design frequency. The useful bandwidth of series-fed collinear antennas is inherently narrow because of the phase error between successive radiating sections that occurs when the frequency is changed from the design frequency. The typical behavior of the major lobes of the vertical radiation pattern of these arrays is shown in Figure 29-10. A further problem with long arrays is that as the array length is increased, the series connection of the elements results in an increased input impedance; as the transformation ratio of the input feed network increases, so the input impedance bandwidth is reduced.

Parallel-Fed Arrays Much greater control is obtained by using an array of fat dipoles with an internal, branched, parallel-connected feed system. Arrays of this type provide stable gain, radiation patterns, and input VSWR over wide bandwidths. A well-optimized eight-element array is able to provide acceptable gain (~10 dBi), radiation patterns, and input VSWR (< 1.7) over the band 225–400 MHz and is suitable for ground-to-air communications; at

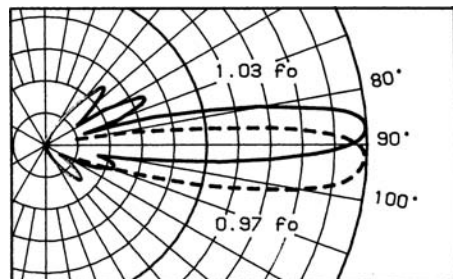


FIGURE 29-10 Vertical radiation pattern of a typical end-fed array

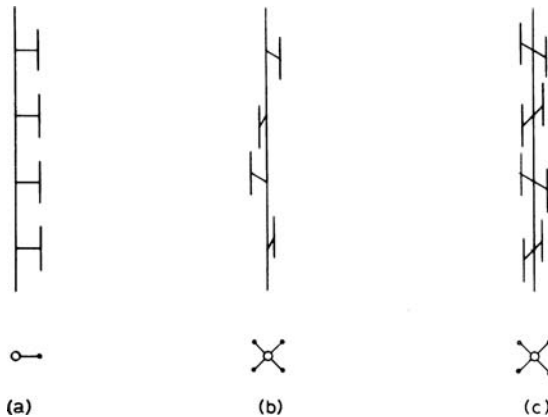


FIGURE 29-11 Pole-mounted dipoles: (a) Inline; (b) Four dipoles spaced around a pole; (c) Eight dipoles spaced around a pole

this frequency the diameter of a structural GRP radome is around 250 mm. Antennas may be designed using a combination of parallel and series feeding; for example, an array can be designed as two end-fed arrays end-to-end, and fed in the center where they join. As requirements for elevation pattern stability increase, the array must be further subdivided into parallel-fed units.

Dipoles on a Pole A simple offset pole-mounted array is shown in Figure 29-11a. This will provide a gain of about 10 dBi in the forward direction but typically only 4 dBi rearward, depending on the pole diameter and the spacing between the dipole and the pole axis. An attempt to avoid this problem is shown in Figure 29-11b, but this type of antenna has distorted vertical radiation patterns caused by the phase shifts that result from the displacement of the dipoles; gain is also reduced to about 6 dBi for the four-element array shown.

The solution in Figure 29-11c, in which dipoles are placed in pairs and are cophased, is more satisfactory, as the phase center of each tier is concentric with the supporting pole. However, the antenna is relatively expensive, as in this arrangement eight dipoles provide only 6-dB gain over a single dipole.

One possibility is to use the inline stacked array in Figure 29-11a and to place the base station toward the edge of the service area. The rearward illumination may be improved if the spacing between the dipoles and the pole is optimized for the pole size and operating frequency.

Analytical solutions to the azimuth pattern are available, and simpler computer programs provide results in good agreement with measurements. When designing the feed networks for multi-element arrays of this type, take care to allow for the effects of mutual impedances, especially when unequal currents or asymmetric geometries are used.

Antennas on the Body of a Tower Figure 29-12a shows a measured horizontal radiation pattern for a simple dipole mounted from one leg of a lattice tower 2-m face width. The distortion of the circular azimuthal pattern of the dipole is very typical and is caused by blocking and reflection from the structure. By contrast, Figure 29-12b shows what can be achieved by an antenna comprising three dipole panels mounted on the same structure. The penalty of adopting this improved solution lies in the cost of the more complex antenna, so before an optimum design can be arrived at, the value of the improved service must be assessed.

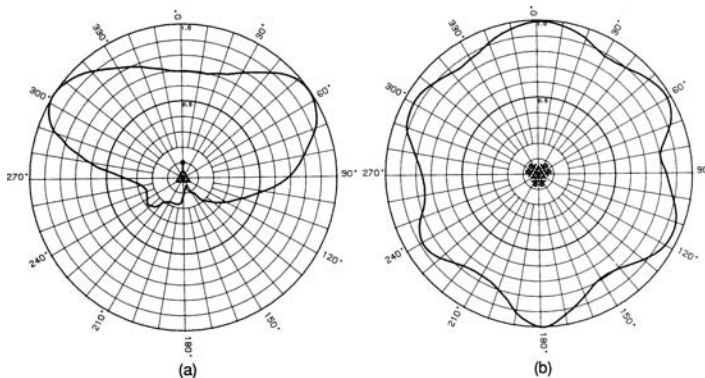


FIGURE 29-12 Typical azimuth patterns of (a) VHF dipole mounted off one leg of a triangular mast and (b) three dipole panels mounted on the same structure (linear field plot)

The horizontal radiation pattern of a complete panel array is usually predicted from measured complex radiation-pattern data for a single panel, using a suitable computer program. For each azimuth bearing, the angle from each panel axis is found, and the relative field in that direction is obtained. The radiated phase is computed from the excitation phases and physical offsets of the phase centers of the individual panels.

Depending on the cross-sectional size of the structure, the most omnidirectional coverage may be produced with all panels driven with the same phase or by a phase rotation around the structure; for example, on a square tower the element current phases would then be 0° , 90° , 180° , and 270° . When phase rotation is used, the individual elements may be offset from the centerlines of the faces of the structure to give a more omnidirectional azimuth pattern, as in Figure 29-13.

A panel array comprising four tiers, each of four panels, is an expensive installation, but if properly designed, it can have a useful bandwidth of as much as 25%. This allows several user services to be combined into the same antenna, each user having access to a very omnidirectional high-gain antenna.

Groups of panels may be arranged and fed to produce an azimuth pattern tailored to the arbitrary requirements of the service area or to provide nulls necessary to meet co-channel protection objectives. Panel arrays are discussed in further detail with reference to their use by broadcasting services in Section 29.9.

VHF/UHF base-station antennas are sometimes situated on the bodies of large towers, perhaps up to 10 m in diameter. It is not economically possible to provide smooth omnidirectional coverage from such a large structure. However, by use of some lateral thinking, it may be possible to

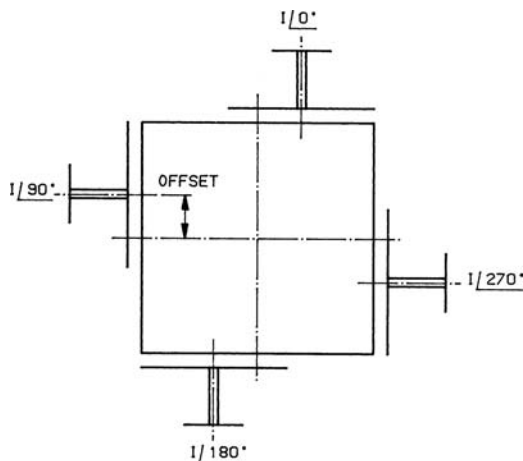


FIGURE 29-13 Plan arrangement of an omnidirectional panel array on a large tower

make a virtue out of a necessity. It is often possible to provide solid coverage of an arc of 270° , or to use two frequency channels, providing overlapping coverage in the region of greatest traffic. If the station operator can guarantee the rearward null in the azimuth patterns at both frequencies, the frequency planner may benefit, and the operator may get two channels where they matter most.

Special-Purpose Arrays For applications in which the largest possible coverage must be obtained, the azimuth radiation pattern of the antenna must be shaped to concentrate the transmitted power in the area to be served, for example, an airway, harbor, or railroad track. Energy radiated in other directions is wasted and is a potential cause of interference to others.

Antennas with cardioidal azimuth radiation patterns are useful for a wide range of applications. Simple two-element arrays (dipole plus passive reflector) or dipoles mounted off the face of a tower may be adequate, but a wider range of patterns is available if two driven dipoles are mounted on a single supporting boom and excited with suitably chosen currents and phases.

When a signal must be laid down over an arbitrarily shaped area of terrain, Yagi-Uda arrays or LPDAs may be arranged as at Figure 29-2*f* and *g*. Due allowance must be made for the separation of the phase centers of the antennas when computing the radiation patterns. As an approximate guide, the phase center of a Yagi-Uda antenna lies one-third of the way along the director array, measured from the driven element. Further tiers of antennas may be used to increase the gain of the system without modifying the azimuth radiation patterns.

29.9 ANTENNAS FOR BROADCAST SERVICES AND BASE STATIONS

One of the characteristics shared by broadcasting antennas and mobile base station antennas is that they are usually linear vertical arrays of some form of standard radiating element. This format derives from the requirement to lay down a signal over some defined area of the surrounding terrain while concentrating the radiated energy into a well-focused beam in the elevation plane; radiation skyward is wasted, so the objective is to concentrate as much energy as possible toward the outer edges of the served area (sometimes, but not necessarily, the horizon). The characteristic of linear arrays is that the radiation patterns obtained in the vertical (elevation) plane and the horizontal (azimuth) plane are mathematically separable. That is, the elevation pattern is determined by the way in which the radiating elements are physically arranged and electrically fed in the vertical plane—the nature of the individual elements only matters to the extent that the pattern is the product of the elevation pattern of the individual elements and the array factor generated by the chosen spacing and complex feed currents. The azimuth pattern is independent of the way in which successive vertical tiers are fed and only depends on the arrangement of a single tier (bay) of elements. It is important to appreciate that this separability collapses if the main beam of the antenna is deflected downward away from the horizontal direction, or if the successive tiers of the antenna do not have identical azimuth patterns. The main difference between broadcasting and base station antennas is that in broadcast practice it is usual for antennas to have wider and more complex azimuth patterns than base station antennas.

Before looking at the methods for the design of these antennas it is worth reviewing some basic array theory. First consider an array of elements spaced in the vertical plane, providing omnidirectional azimuth coverage. If the aperture is large, say more than eight wavelengths, the total directivity is only a function of the length of the array. We can fill it

with elements with low directivity (isotropic radiators) or higher directivity (for example, vertical dipoles one wavelength long), but as long as we *fill* the aperture with elements, the directivity will be the same. (This doesn't apply in the same way to short arrays because the effective aperture can "overhang" the physical aperture by a more significant proportion.) As a rule of thumb, the maximum directivity of an omnidirectional array will be about 1.15 times per wavelength of vertical aperture. If the aperture is underfilled—the elements are too far apart—the result will be lower directivity because energy is lost from the main beam into characteristic high elevation sidelobes (grating lobes). Overfilling the aperture with elements produces no benefit in directivity and increases the coupling (mutual impedance) between adjacent elements, making it more difficult to control the excitation of the array.

In the azimuth plane the situation is different. If we select a radiating element whose property is to uniformly illuminate a 180° sector with no signal at all in the other half-space, then it is clear that such an element has a directivity of two (3 dB). Because directivity is properly defined as a 3D property, this ratio is commonly known as the *azimuth max-to-mean ratio*. It is numerically equal to the area of the circle circumscribing the polar radiation pattern when plotted in linear field coordinates, $E(\theta)$, to the area of the plotted radiation pattern itself (because both power and area are proportional to E^2). This makes it very easy to visualize what is happening.

The maximum total directivity of a vertical array is the product of the directivity in the elevation plane times the azimuth max/mean ratio.

Azimuth Pattern Shaping Each tier of an array may comprise a single radiating element; this is common in base station practice and for simple broadcast antennas. The desired pattern is often obtained by arraying elements against a mounting pole, or against a reflecting plane when more suppression of radiation is needed in the rearward direction. Where the required azimuth pattern cannot be achieved using a single radiating element (with or without a reflector) each tier is constructed from several elements, each often constructed in the form of a panel comprising an integrated radiating structure and reflector. Panels are usually designed to cover wide frequency bandwidths and to maintain very consistent radiation pattern, polarization, and impedance over the operating band. Typical panels cover entire broadcast bands. Designing arrays in this manner has the advantage that by using standard, highly optimized, mass-produced panels the antenna designer can construct a wide variety of azimuth pattern shapes. The adoption of sector coverage for mobile radio base stations has removed the requirement for separate panels, but the design of the radiating element has the same constraint of requiring constant performance over wide bandwidths. To reduce the mutual impedance between adjacent elements individual elements are often surrounded by fences or cavities.

Elevation Pattern Shaping

There is little point in directing the elevation maximum of our vertical array in exactly the horizontal plane: in flat terrain half the radiated power will be lost over the head of any ground-based receiver. The further we get from the antenna the more the curve of the earth will result in the main beam passing overhead. The main beam is therefore usually tilted downward, typically so the upper -3 -dB point lies in the horizontal plane; the exact tilt may depend on the relative elevation of the antenna and the served users. A further result of tilting the beam downward is to reduce the level of unwanted illumination of users located beyond the effective served area. The served area is always surrounded by a region covered by signals too weak to use but too strong to ignore; the size of this zone is reduced when beamtilt is used. For ground-to-air services a small upward beamtilt is sometimes applied.

A column of elements with identical currents in each radiating element creates an elevation radiation pattern with a simple $(\sin x)/x$ form. This pattern is generally modified to create some additional shaping by feeding the tiers of the array with currents of different phases and/or amplitudes. There is no unique solution to the creation of a radiation pattern with some arbitrary shaping (as we generally need to match only a requirement for amplitude and not for both amplitude and phase), so a variety of procedures can be adopted to derive a suitable set of element currents. Practical considerations include ensuring that the chosen function does not require very different currents in adjacent elements, which will exaggerate the effects of mutual impedances. The radiation pattern must not be sensitive to small changes in some of the array currents, as this may make it difficult to achieve the required pattern over a wide frequency band.

A simple method of filling nulls and suppressing sidelobes is based on the principle of superposition. Given that a uniform array has an elevation pattern in the form of a sinc function, supplementary excitations can be added to the principal excitation with chosen amplitudes and relative phases. To fill a null we apply a sinc function with its maximum directed at the angle of the null and with its phase in quadrature with the principal excitation; to suppress a sidelobe an additional sinc function is added with its maximum in the direction of the sidelobe and in antiphase with the main excitation. Additional sinc functions can be added to fill successive nulls or to suppress as many sidelobes as required. There is some interaction between the added sinc functions because each has its own sidelobes at -13 dB relative to its maximum, but by using a computer program the method can be implemented in an interactive manner. At the end of the optimization process the element currents associated with each sinc function are added vectorially to give the complex current that will be needed to generate the specified pattern.

When assessing the performance of an array it is useful to obtain a sample of the radiating currents from each element by using inline couplers or external sampling loops. The currents can be displayed in polar coordinates using a vector network analyzer to allow the designer to verify that the actual currents are close to the design currents and to maintain the correct relationship within some acceptable limit over the operating frequency band. The sensitivity of the radiation pattern to small variations in the element currents can be investigated using a Monte Carlo analysis in which random variations in amplitude and phase, with a known standard deviation, are applied to the array currents, and the resulting patterns are evaluated for compliance with the target specification. The most common reason for element currents deviating from their design values is the imperfect impedance match of the elements; this causes the power division at each junction to deviate from the expected value in a manner that is often strongly frequency dependent. The design of many radiating elements in broadcast and base station arrays is often governed by the requirement that individual elements (or element groups) have a sufficiently low VSWR to allow radiation pattern constraints to be met over extended bandwidths. The use of Wilkinson power dividers can limit the extent of current variation, but unless good matching is achieved this may be at a price of losing RF power into balancing loads.

Impedance Characteristics The maximum permissible input VSWR of an antenna is determined by the effect that a significant reflection would have on the transmitted signal. The applicable criterion is commonly more stringent than considerations of transmitter currents and voltages or of maximum power transfer to the antenna. With respect to the input impedance of the whole array, there are several important considerations:

The radiation of signals with relative time delays If we examine the typical arrangement shown in Figure 29-14, we see that a transmitted pulse, which may be part of any analog or digital transmission, travels from the transmitter to the antenna, experiencing some loss and transit time in the interconnecting cable. If the antenna has a return loss L db,

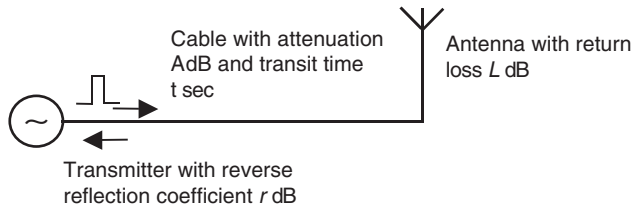


FIGURE 29-14 Transmitter feeding a mismatched load

a signal L dB below the forward signal at the antenna now moves along the cable in the reverse direction. After suffering a delay t and attenuation A , the signal encounters the output port of the transmitter. In general the impedance looking back into a transmitter is very far from $50\ \Omega$, so much of the signal is now reflected with reflection coefficient r dB. After again passing along the cable, the signal originally reflected at the antenna is reradiated with a relative level of $-(L + 2A + r)$ dB and a delay of $2t$ relative to the original signal. The significance of both the amplitude and the relative delay of this secondary signal depends on the signal format being transmitted. In analog TV and digital systems using high-order digital modulation formats it is the essential determinant of the required antenna reflection performance, especially for antennas fed by long, low-loss cables.

Frequency-dependent load impedance at the transmitter If the antenna input impedance differs from the characteristic impedance of the main feeder, the impedance seen by the transmitter varies with frequency. (The impedance of the antenna forms circles round the center of the Smith Chart.) The result of a transmitter working into an impedance that varies with frequency is often to distort the output waveforms in both phase and amplitude. Again the effect and its acceptable limits depend on the signal format. With some systems it is usual to adjust the transmitter characteristics to compensate for some degree of distortion caused by the varying load impedance, but the extent to which this can be done is limited, especially when the antenna is connected by a long, low-loss cable.

Cable power ratings The effective power rating of a cable is reduced when it is operating at a high VSWR, as the high-current points in the standing wave cause hotspots on the cable; for this reason the cable rating must be reduced by a factor of $1/\text{VSWR}$.

These effects usually define the required impedance characteristic of the whole antenna. The effect of reflection loss only becomes significant at high values of VSWR, the rule-of-thumb values being 0.5 dB at 2:1 and 1 dB at 2.6:1 (return losses of 9 dB and 7 dB, respectively).

There is a further reason why the VSWR of individual elements needs to be controlled. If two loads are connected in parallel at a transmission line junction, the current division at the junction is in inverse ratio to the (complex) impedances seen at the inputs to the separate branches. If we join two mismatched antennas to a junction through cables of *equal* length, this presents no particular problem. Although mismatched, the impedances are equal so the current divides equally in amplitude and phase. If the lines connecting the loads to the junction are of *unequal* length—perhaps because we require a phase shift between the currents in the connected elements—the impedances presented at the junction are no longer equal because they lie at different points on a circle round the center of the Smith Chart. The result is that the load currents will be neither equal nor of the expected relative phase, modifying the radiation pattern of the arrayed elements in a manner we may not have expected. This imposes a new criterion for the input match of the radiating

elements, depending on the accuracy with which design currents must be achieved and maintained over extended bandwidths in order to create the required radiation patterns. If the equality of the currents must be preserved despite mismatched impedances, the use of a hybrid junction or Wilkinson splitter will reduce the unwanted effect of the mismatch on the power division.

In situations in which antenna matching is very critical it is common to use a variable transmission line device to center the complex impedance plot of the antenna relative to the exact Z_0 of the coaxial feedline. The variable components may take the form of short sleeves (“slugs”) fitted to the inner conductor of a section of transmission line—the equivalent of small patches on a microstrip line. Where close impedance matching is needed at a number of different frequencies, a long section of transmission line is used and is fitted with several slugs at different computed positions.

Designing Radiating Panels

The commonest designs for radiating panels provide an azimuth beamwidth of 90° (total) at -6 dB relative to boresight. When two panels driven with equal co-phased currents are mounted close to one another on the sides of a square tower, the field in the diagonal direction is equal to the field in the main beam direction of each panel. A successful panel must have an azimuth beamwidth that is very stable with frequency, and the level of any subsidiary lobes must be very low. Typical panels comprise two full-wave dipoles mounted against a reflecting screen, usually made from solid sheet above about 400 MHz and formed from conducting rods at lower frequencies. Because antenna installers use horizontal bars as handholds and climbing steps, it is common to make these bars from hot-galvanized mild steel. In the UHF band the design and production of wideband panels with well-controlled azimuth beamwidth and low VSWR (typically $< 1.1:1$ over 470–860 MHz) presents an interesting challenge.

Combining Panels into Groups The most usual feed method is to connect all the panels that form one tier of an array together into one power divider, constructed from coaxial line sections for use at high powers or various forms of microstrip for low powers. The input ports of each tier are then combined, usually in a branched tree pattern. For base station antennas a single feed is usually provided to the whole array, while in broadcast practice it is usual to provide separate inputs to each (vertical) half of the whole array, and to connect the antenna to the final stage of combining at ground level. This configuration gives access to the two separate half-antennas, allowing some redundancy in the case of failure of some part of the array, and access for maintenance when one half array is disconnected and grounded—in these cases the gain of the antenna is obviously reduced when only one half is driven, but loss of some coverage is better than no transmission.

It is obviously also possible to combine a broadcast array by interconnecting all the panels on each face with a power division network and then to interconnect the faces in a final stage of combining. This method is not favored because it uses more cable, and a single fault may grossly modify the azimuth radiation pattern. When multiple radiating elements on each panel are used to create circular polarization it is usual to combine the elements of each panel together before feeding them from the main feed network.

For analog TV transmissions it is necessary to achieve a very low VSWR across each transmission channel, and arrays often have extremely wide bandwidths. For this reason the phase rotation method shown in Figure 29-15 has been extensively used. It can be extended by feeding adjacent tiers in phase quadrature (and rotating them mechanically by 90°) and repeating this with successive groups of two and four tiers. Complete half antennas are often fed in phase quadrature without a corresponding mechanical rotation to

provide elevation beamtilt equal to half the total elevation beamwidth. These methods are not so easily applied when the azimuth radiation pattern is directional.

Directional radiation patterns are achieved by a combination of several methods:

- Adjustment of the amplitude and phase of panels forming each tier of the array
- Adjustment of the physical positions of the panels in each tier
- Use of unequal number of panels on different faces of the array (in this case the physical configurations of different tiers are not identical)

Azimuth pattern requirements are driven by a mixture of the minimization of interference with stations operating on the same or adjacent channels (which may be contained in international regulatory requirements), topography, and the distribution of the target population. Azimuth patterns can be predicted by computer programs (which often make use of measured data from individual panels) but usually need validation by practical measurement—at least of an appropriate subsection of the complete array. The most common sources of difference between predictions and measurements occur when:

- The array has substantial physical displacement of panels—potentially allowing blockage of one panel by another.
- Current paths in the screening surfaces are physically interrupted; such a problem occurs in a horizontally polarized array if the screening panels are not bonded to one another at each corner, when the gaps at each corner operate as slot radiators that were not included in the computer prediction.
- Radiating elements with imperfect matching are connected together by cables of unequal length (see above).

High-power broadcast arrays require very careful attention to mechanical design details, and the advice in Table 29-1 needs to be thoroughly applied. The accumulation of ice and snow will change element matching with the result that radiation patterns, gain, and input VSWR will all be degraded. High voltages at element feedpoints are likely to create dielectric heating, and any insulators are likely to suffer surface tracking both from exposure to the weather and accumulation of contaminants from urban rain, birds, and salt spray at coastal locations. This problem can be addressed by the use of a radome, fitted either over the whole array or over the elements or their feedpoints—in this last case the radome must be sufficiently large that it does not experience strong surface fields, or the result will be failure by surface tracking.

Omnidirectional Broadcasting Antennas Broadband high-power omnidirectional antennas are usually built from tiers of four (occasionally three) panels that may have linear or circular polarization. Depending on the cross-sectional size of the structure, the most uniform omnidirectional coverage may be produced with all panels driven with the same phase or by a phase rotation around the structure; for example, on a square tower the element current phases would then be 0° , 90° , 180° , and 270° . When phase rotation is used, the individual elements may be offset from the centerlines of the faces of the structure to give a more omnidirectional azimuth pattern, as in Figure 29-13.

Omnidirectional vertically polarized antennas can be produced by mounting two vertical dipoles either side of a supporting pole, carefully choosing the spacing to provide a uniform azimuth pattern. If the support is too large (in wavelengths), then three or more dipoles may be needed to provide uniformity.

Horizontally polarized omni antennas can be constructed using a *turnstile* format in which two separate elements are arranged at right angles and fed in phase quadrature.

An alternative method is to create a *square loop* from four horizontal half-wave dipoles fed in phase. This produces a good omnidirectional azimuth pattern but the large capacitive coupling across the corners of the array limits the achievable impedance bandwidth (see Figure 29-15a). Circularly polarized antennas are sometimes based on loop-dipole derivatives and short radial-firing multiple-start helices. These are often mounted off the side of the supporting structure, providing quasi-omni patterns with degraded polarization circularity.

An omnidirectional array can be designed using broadband panels with crossed-dipole elements, but this arrangement is costly and has a very substantial weight and windload. The pole-mounted Lindenblad array²⁰ (see Figure 29-15b) is formed from four half-wave dipoles mounted on a square that is a half-wavelength across, with the elements lying at 45° to the horizontal and driven with equal cophased currents. When viewed from the side of the square, it is easily seen that the geometry ensures that the vertical and horizontal field components are equal in magnitude and have a quarter wavelength relative to the physical displacement of their phase centers. The arrangement provides an omnidirectional azimuth pattern with low ripple and good polarization purity. An array comprising multiple tiers, centrally mounted on a supporting pole, can provide high gain and low VSWR across the whole FM radio band 97.5–108 MHz and is suitable for high-power multi-channel operation.

A short four-start helix can be imagined as a derivative of a Lindenblad antenna using curves radiating elements. A modified form of Lindenblad array is sometimes used in which the four dipoles are fed with phase rotation, the spacing between dipoles reduced, and the tilt angle changed to around 30° . This provides hemispherical coverage but is not suitable for terrestrial broadcast use.

Feeding Multiple Arrays Simple low-power arrays operating at frequencies below 500 MHz are usually fed using a coaxial-cable branching network (see Figure 29-16). Stripline power dividers are attractive for applications above about 500 MHz and can be designed to provide arbitrary power division ratios and any required number of ports; they provide a high level of reproducibility in volume production and can be designed using readily available computer software. If an array is to operate at high mean input power, it may be necessary to use large-diameter fabricated coaxial transformers.

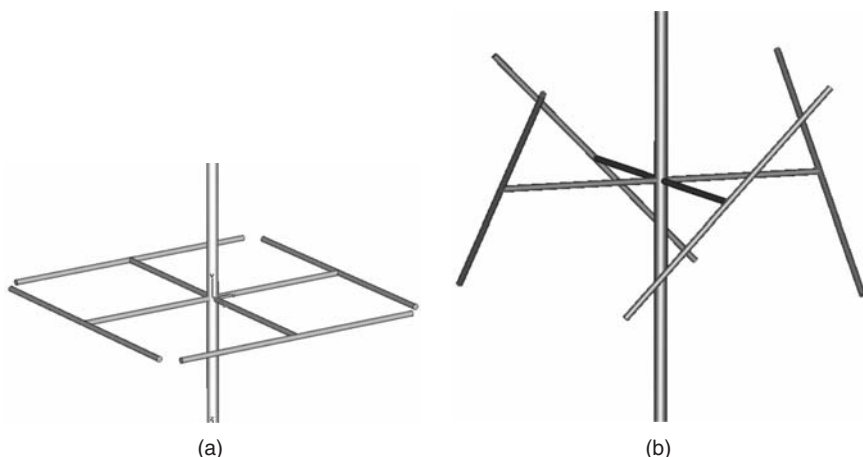


FIGURE 29-15 (a) A horizontally polarized square loop comprising four half-wave dipoles; (b) A circularly polarized Lindenblad array in which the dipoles are rotated by 45°

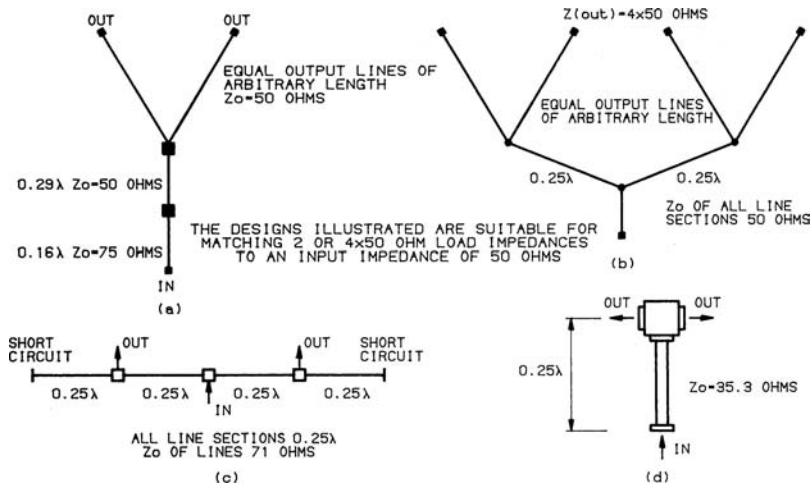


FIGURE 29-16 Simple branching feeder systems: (a) Two-way; (b) Four-way; (c) Two-way, compensated; (d) Two-way, high-power

Multiway broadband power dividers are usually designed using multiple quarter-wave sections to achieve a very low input reflection coefficient over a very wide bandwidth. The general principle used is illustrated in Figure 29-17. At the common point of the branching output lines the impedance is low and requires to be transformed up to 50 Ω . If this is done in a single step over a wide bandwidth, the transformed impedance will take the form of an arc on the Smith Chart as in Figure 29-17a. A first transformation is made to an intermediate point on the Smith Chart, and this is followed by a second transformation with a similar impedance ratio. The action of the second quarter-wavelength is again to wind the intermediate impedance into an arc in which the angle traversed is proportional to frequency; this compensates the behavior of the first transformation, and the result is a very small impedance plot lying close to 50 Ω (see Figure 29-17b). When large numbers of branch lines are to be fed it is good practice to transform the outgoing line impedances up before connecting them in parallel, to avoid the need to make lines with extremely low Z_o . Where large power division ratios are needed the outgoing impedances can be transformed up so

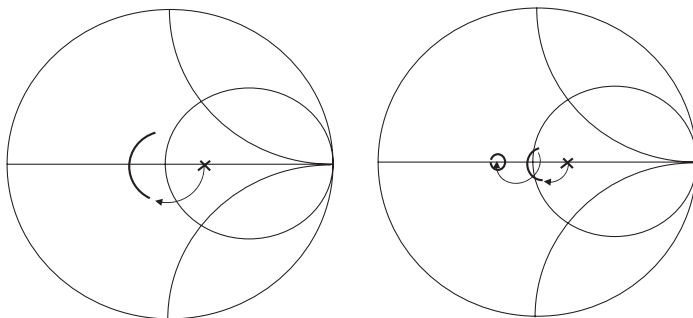


FIGURE 29-17 (a) The dispersive effect of a simple quarter-wave section over a wide bandwidth. (b) A two-stage transformation has been used in which the effect of the second section is to reduce the dispersion caused by the first section.

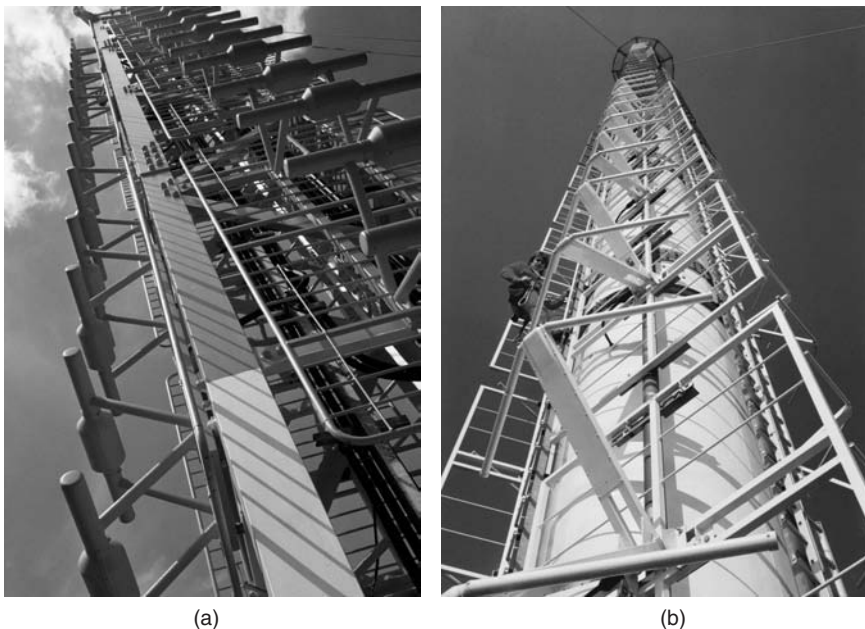


FIGURE 29-18 (a) Panel arrays of full-wave dipoles for 174–230 MHz and (b) bent half-wave dipoles for 87.5–108 MHz. The rigger is a reminder of the need to plan for safe access and to ensure that equipment cannot be damaged by being used as footsteps or handles. (Courtesy of Kathrein-Werke KG)

the impedances at the common point have the required ratio and are stable with frequency; this ensures that the power division ratio is stable. Designs of this type can be realized in coaxial or microstrip lines.^{21–24}

Alternative designs can be created by thinking in lumped element terms and using short lines of high Z_0 (inductors) and of very low Z_0 (capacitors). This technique can produce some very compact splitters with wide frequency bandwidths. Wilkinson power dividers have characteristics similar to those of a hybrid junction.

Figures 29-18 and 29-19 show typical examples of high power radio and TV broadcast antennas embodying many of the techniques that have been discussed.

29.10 BASE STATION ANTENNAS FOR MOBILE PHONE SERVICES

In many respects the design of base station antennas parallels that of a directional broadcast array. In both systems the shaping of the azimuth radiation patterns is used to control coverage and co-channel interference in the coverage area of other stations. Both systems employ columnar arrays of radiating elements in which the currents in different tiers are closely controlled in order to shape the elevation radiation pattern—elevation null-fill is used to avoid holes in local coverage. The main differences are the azimuth pattern (base stations have radiating elements firing in only one direction), transmitted power (a few hundred watts rather than tens of kilowatts), and the scale of production—these differences have led to the widespread use of microstrip techniques in both radiating elements and feed systems.

The use of circular polarization in many broadcast systems is paralleled by the use of polarization diversity in base station antennas (transmitting independent linear components rather than combining them in phase quadrature). Broadcast antenna engineers will find many interesting techniques in base station antenna design, where the changing demands of customers and intense competition have fostered rapid development of simple but effective broadband methods adapted to large-scale production.

The parameters of a base station antenna are an essential means by which the coverage and capacity of the network are optimized. Space-diversity receive systems employing two vertically polarized antennas are common in rural areas; in urban areas polarization diversity is often used and is provided by antennas supporting two orthogonal linear polarizations ($\pm 45^\circ$). Because of the closely spaced channels used by a base station, the transmitters are often combined using 3-dB hybrids (incurring a 3-dB loss). To minimize combining losses, all available antenna ports are used for multichannel transmission as well as for reception, so there are very stringent requirements on passive intermodulation products. The development of an increasing number of frequency bands for mobile radio has led to the development of dual-band antennas (usually 850/1900 MHz or 900/1800 MHz); the advent of 3G services in the 2100-MHz band has encouraged the development of wideband antennas operating over the ranges 826–960 MHz and 1710–2170 MHz.

Further developments in base station antennas include the use of remotely adjustable beamtilt, azimuth beamwidth, and azimuth pointing direction, and the integration of low-noise amplifiers into antenna systems. The continuing pressure to meet rising capacity and coverage needs has led to the development of inconspicuous “street-works” base stations with three sector antennas, each with remote tilt facilities, housed in a single radome and mounted on a street-lighting pole.

A **vertically polarized antenna** comprises an array of vertically polarized dipoles mounted in front of a reflecting screen and protected by a radome, usually made from UV-stabilized ABS or polystyrene. Successful designs have used

- Punched or pressed dipoles fed by miniature coaxial cables, using cable or microstrip power dividers
- Microstrip dipoles with a distribution network integrated in a single PCB or fabricated assembly
- Dipoles fed by suspended microstrip feeds, all cut from aluminum sheet
- Stacked patches with microstrip feed networks



FIGURE 29-19 Typical broadcast antenna installations on a concrete supporting structure. A UHF TV antenna is mounted on top of the tower. It is protected by a GRP radome that is provided with helical strakes to reduce its vulnerability to mechanical oscillation at certain wind speeds. Below this is a VHF FM radio antenna built from dipole panels, and below this is a variety of LPDAs. (Courtesy of Kathrein-Werke KG)

Dual-polar antennas have been produced using

- Crossed dipoles constructed by printed-circuit or sheet metal techniques, integrated with microstrip or cable feed systems.
- Squares comprising four half-wave dipoles, usually of stamped or cast metal construction with cable feed.
- Patch elements, usually air-spaced and standing 10–20 mm above the reflecting plane, fed as stacked patches or by some other indirect feed mechanism below them. (Directly edge-feeding a wideband patch gives rise to pattern squint at frequencies far from the design frequency.)

Dual-polar antennas require the specification of several additional parameters to ensure optimum operation in a mobile radio system:

Azimuth radiation patterns and gain are defined assuming the antenna under test is illuminated by a plane-polar signal with a polarization of either $+45^\circ$ or -45° . Because the sense of slant polarization in the rearward direction is opposite to that at 0° , the cross-polar rear lobe often limits network performance and should always be measured. Some networks define a sum-power limit, but this is pessimistic as no real antenna could receive the sum power in both polarizations. The azimuth pattern is defined and measured as a conical cut of the 3D solid pattern, with a half-angle equal to the complement of the nominal elevation beamtilt. Errors in the azimuth pattern relate to the design of the radiating elements and their immediate environment in the array.

Elevation beamtilt is conventionally defined in the sense that positive beamtilt is downward. The specified tilt must be achieved within some defined limit that must be maintained over the whole operating bandwidth. Changes in tilt result in uncertainty in the coverage footprint of the base station between transmit and receive frequencies and variable overlap with adjacent stations. The impedance matching of the upper and lower halves of the array is the first matter to investigate if an array has an unstable beamtilt; any significant mismatch will cause a frequency-dependent phase error between them.

Squint is the angle between the azimuth directions of the beams formed in the two polarizations, usually measured as the mean direction between the -3 -dB points of the azimuth patterns in each polarization. Squint is controlled by the electrical symmetry of the radiating elements and their means of excitation. It is usually greatest for large tilt angles because the use of $\pm 45^\circ$ polarization puts any element squint into the diagonal planes: the larger the elevation tilt, the more any element squint expresses itself in the azimuth pattern of the array.

Tracking is the gain difference between the two polarizations measured over some specified range of azimuth angles surrounding the main beam direction, so this parameter includes the effects of both squint and beam-center gain difference. If there are significant frequency-dependent tracking errors, a mobile may be handed-off into the cell on the basis of a strong signal in the broadcast control channel (BCCH), but if the polarization tracking is poor and the assigned channel is transmitted in the other polarization from the BCCH, the signal may be several dB below what would have been expected. Precise tracking is achieved by minimizing squint and making the $+45^\circ$ and -45° subsystems as mechanically and electrically equal as possible.

Cross-polar or interport isolation (XPI or IPI) is the attenuation measured between the $+45^\circ$ port and the -45° port. This is significant because of the practice of using both ports for transmission—the coupled energy between transmitters must lie below -30 dB because this is the level at which the transmitters meet their spurious emission specifications. This is a stringent requirement and various coupling compensation mechanisms are used, including tabs on the elements, rods, bars, and “mushrooms” placed between elements and along the edges of the antenna chassis, metal tapes inside radomes, and feedline coupling. The XPI usually increases with increasing elevation tilt because the phase shift across the array results in the coupled signals being increasingly out of phase.

Cross-polar discrimination (XPD) is the relative response of one section of the antenna to co- and cross-polar signals. A more logical parameter to define is *polarization orthogonality*, which is the parameter that permits the antenna to provide polarization diversity, but this is mathematically complex and is not easy to measure. It is a common feature of directional slant-polar antennas that their polarization on boresight is close to 45°, but as the observer moves away from boresight the polarization tends to become dominated by the vertical component (at 90° from boresight it is easy to understand that the polarization becomes completely vertical). In practice, the diversity gain obtained from a dual polar antenna does not fall significantly for values of orthogonality or XPD larger than about 10 dB,²⁵ so this parameter is *not* often defined or measured. The achievement of a large value of XPI does not ensure that the XPD is high, and the XPD should always be checked, especially if large amounts of compensation are necessary to achieve the required XPI.

A 45°-slant polarization has the unusual property that if the transmitting antenna has +45° polarization, the polarization appears as -45° to the antenna that receives it. A base station antenna is used for both transmission and reception, and there is no clear convention as to how ports should be labeled. If they are used or measured in any manner in which the distinction is relevant, the port naming should be regarded as arbitrary and the actual polarization checked.

Cross-band isolation is defined for multiband antennas to ensure that the signals presented to the base station filter system lie within the expected limits. In antenna terms, insufficient cross-band isolation is likely to indicate that currents in the high-band part of the antenna will excite radiating currents in the low-band elements, corrupting the high-band radiation patterns. If an antenna has adequate radiation patterns but some deficit in cross-band isolation, it is acceptable to add filtering in the feed system.

Some antennas are required to provide a single input port for both frequency band groups, in which case a diplexer will be needed to combine the feed systems of the high- and low-band arrays. This can be formed from high-pass and low-pass elements engineered in microstrip and needs sufficient isolation to prevent the radiation pattern distortion explained earlier.

Passive intermodulation products (PIMs) of all orders are generally required to lie below -153 dBc for two carriers at a level of +43 dBm (20W). For dual band antennas the first-order transmit-band products may lie in the receiver band and can be particularly troublesome ($900 + 900 = 1800$; $1800 - 900 = 900$). Good PIM performance is only achieved by the exercise of good design practices and of care and cleanliness in manufacture (see Section 29.2).

Input VSWR is usually required to be less than 1.4:1 across the relevant operating bands. It is usually most difficult to achieve this if the array has a 0° electrical beamtilt because all the elements are driven in phase so the reflections also arrive back in phase at the input. As the beamtilt is increased the reflections are no longer cophased, so the required VSWR is easier to achieve.

Some design margin is needed on these parameters between the specification and what is measured on prototypes and pilot production, in order to accommodate production variation without products failing to meet specification. In this respect PIM, XPI, and VSWR are the parameters most likely to vary significantly between samples; variation is controlled by adopting good design practices, ensuring components are accurately made, and educating production personnel about consistency and cleanliness.

Adjustable Elevation Tilt As noted earlier, the use of beamtilt has long been a feature of both base station and broadcasting antennas. The use of remotely controlled adjustable tilt for base station antennas was introduced as a method by which the network operator can control the area covered by a base station antenna. Reducing the transmitted power also reduces coverage in buildings and shadow areas close to the base station, while down-tilting the main beam allows close control of the position and depth of the overlap area

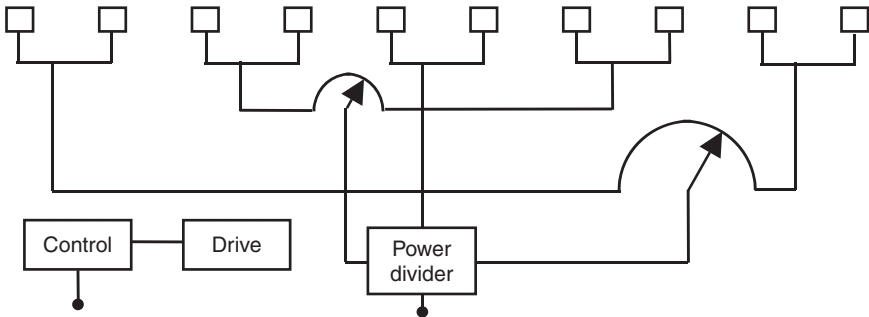


FIGURE 29-20 A typical array with adjustable electrical tilt, fed as five subarrays, each of two elements, using tapped-line phase shifters powered from a single drive motor. Feedlines are usually dimensioned so that when the phase shifters are at one end of their travel, the tilt is 0° .

between base stations. Electrical downtilt (produced by introducing a progressive phase shift over the aperture of the antenna) is used in preference to mechanical tilting because mechanical tilt changes the shape of the footprint as well as its size, and physically downtilted antennas are visually more obtrusive than untilted antennas. The requirement for very low PIM levels and substantial transmitted power has led to the development of electro-mechanical phase shifters driven by stepper motors. Three different forms of design are in general use: transmission lines whose electrical length is changed with moving dielectric loading, lines with capacitively coupled moving taps, and lines with trombone sections. In applying these to a linear array it is necessary to arrange that the phase shift for each element is proportional to its distance along the array. To reduce the total phase shift needed it is common to use a differential scheme in which, to tilt the beam down, elements at the bottom of the array are delayed in phase and simultaneously those at the top are advanced. To reduce the complexity of the phase shifters, the elements of the array are often grouped together into fixed subarrays of two or three elements, and variable phase shift is applied between them (see Figure 29-20).

Grouping elements into subarrays to some extent compromises the performance of the array; a truly linear phase shift produces lower elevation sidelobe levels even at extreme tilts (see Figure 29-21). An industry-standard digital control interface has been defined and implemented by all major antenna manufacturers—this interface²⁶ provides control for antennas, tower-mounted amplifiers, and other tower-top hardware.

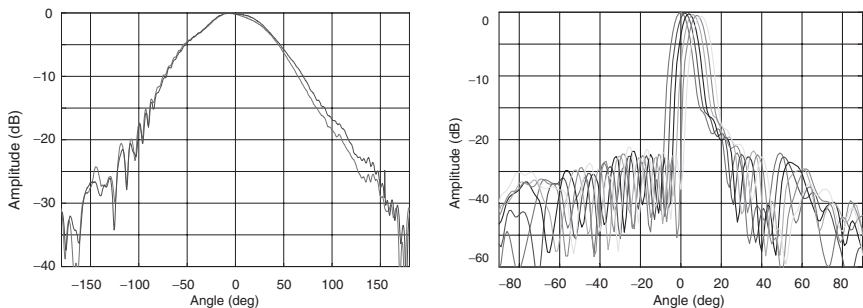


FIGURE 29-21 Examples of the azimuth (left) and elevation patterns (right) of a base station antenna showing excellent consistency of beam shape and sidelobe levels over a range of electrical elevation beam tilt angles of 0° – 10° (Courtesy of Argus Technologies)

Extension of remote control techniques to the adjustment of azimuth pattern characteristics—beamwidth and pointing direction—is relatively easy and provides increased adaptive control of network optimization, allowing real-time response to changing demands for capacity and coverage in the network. Such techniques are based on cross-layer communication originating from radio resource management; they are symmetrical in their enhancement of uplink and downlink and are protocol independent. Antenna systems embodying them have been described as *semi-smart antennas*.²⁷

29.11 SYSTEM CONSIDERATIONS

Mounting Arrangements When mounting any antenna it is important not to impair its performance by the influence of the supporting structure. The inevitable effect of the supporting structure on the radiation pattern of a dipole has been referred to in Section 29.8. This effect is accompanied by a modification of the input impedance, which may be unwelcome if a low VSWR is needed. In any critical application the change of the radiation patterns and gain must be taken into account when estimating system performance. Impedance matching of the antenna must be undertaken in the final mounting position or a close simulation of it.

If Yagi-Uda arrays are mounted with their elements close to a conducting structure, they too will suffer changes of radiation patterns and impedance. The effects will be greatest if tower members pass through the antenna, as they do when an array is mounted on clamps fitted at the center of the cross boom. If at all possible, when an array is center-mounted, the member to which it is clamped should lie at right angles to the elements of the array.

Currents induced in diagonal members of the supporting structure will cause reradiation in polarization planes other than that intended. This will result in the cross-polar discrimination of the antenna system being reduced from that which would be measured on an isolated antenna at a test range. When polarization protection is important, the tower should be screened from the field radiated by the antenna with a cage of bars spaced not more than $\lambda/10$ apart, lying in the plane of polarization. (A square mesh is used for circular polarization.) Panel antennas are designed with an integral screen to reduce coupling to the mounting structure.

Long end-mounted antennas are subjected to large bending forces and turning moments at their support points. These forces can be reduced by staying the antenna, using nylon or polyester ropes for the purpose to avoid degrading its electrical characteristics.

In severe environments antennas may be provided with radomes or protective paints. It is very important that the antennas be tested and set up with these measures already applied, especially if the operating frequency is in the UHF band.

Coupling A further consideration when planning a new antenna installation on an existing structure is the coupling that will exist between different antennas. When a transmitting antenna is mounted close to a receiving antenna, problems that can arise include

- Radiation of spurious signals (including broadband noise) from the transmitter
- Blocking or desensitization of the receiver
- Generation of cross-modulation effects by the receiver

The last two effects depend critically on the isolation between the antennas and on parameters of the transmitters and receivers; these parameters are generally specified by their manufacturers.

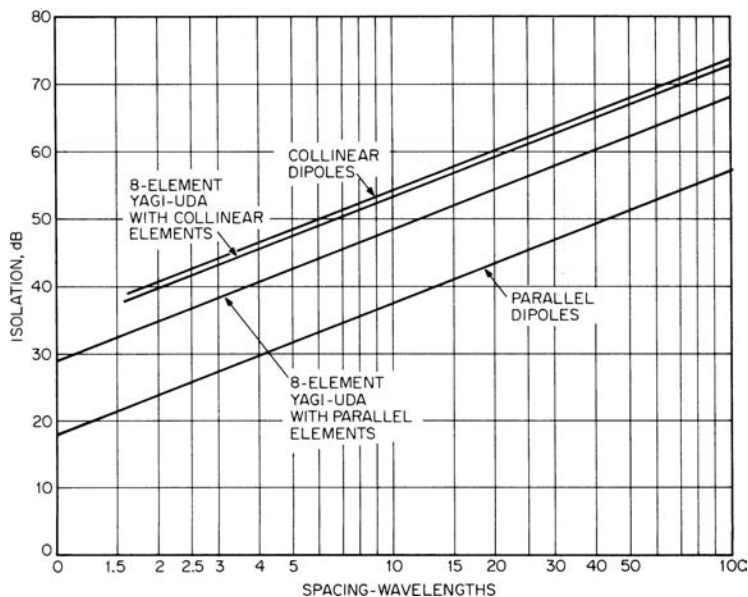


FIGURE 29-22 Typical isolations between Yagi-Uda antennas

The isolation between two antennas may be predicted from Figure 29-22, measurements, or from standard propagation formulas. Antenna isolations may be increased by using larger spacings between them, arranging that they are cross-polarized, or by using arrays of two or more antennas spaced to provide each with a radiation-pattern null in the direction of the other.

An alternative method of increasing the isolation between the antenna-system inputs is to insert filters. If a suitable filter can be constructed, the antenna isolation may be reduced until, in the limit, a single antenna is used with all equipment, transmitters, and receivers coupled to it through filters. When receivers are connected to a common antenna, the signal from the antenna is usually amplified before being divided by a hybrid network. The number of services that use a single antenna can be extended to six or more, provided adequate spacings are maintained between the frequencies allocated to different users. The whole system is expensive, but the cost may be justified if the antenna itself is large or if tower space is limited.

Multiple User Systems To reduce the environmental impact and cost of antenna systems, users share antenna resources by installing a single broadband antenna system and using it to support multiple radio transmitters or receivers (or both). Not only does this practice reduce the profile of an antenna installation, but it also allows several services to make best use of locations where antennas can be most favorably sited.

REFERENCES

1. E. Simiu and R. H. Scanlan, *Wind Effects on Structures: An Introduction to Wind Engineering* (New York: John Wiley, 1986).
2. H. H. Uhlig, *Corrosion and Corrosion Control* (New York: John Wiley, 1985).

3. M. G. Fontana, *Corrosion Engineering*, 3rd Ed. (New York: McGraw-Hill, 1986).
4. *Lightning Protection Code*, ANSI C5.1-1969 (New York: American National Standards Institute, 1969).
5. *Code of Practice for the Protection of Structures Against Lightning*, BS6651:1990 (London: British Standards Institution, 1990).
6. R. H. Golde (ed.), *Lightning*, Vol. 2: *Lightning Protection* (New York: Academic Press, 1977).
7. "Electromagnetic Compatibility and Radio Spectrum Matters (ERM); Radio Site Engineering for Radio Equipment and Systems," *ETSI Guide, ETSI EG 200 053*, (Sophia Antipolis, France: European Telecommunications Standards Institute, 2004). This is an excellent general guide to good site practice, available from www.etsi.org.
8. On the restriction of the use of certain hazardous substances in electrical and electronic equipment, *Directive 2002/95/EC of the European Parliament* (and documents cited therein).
9. *Guidelines for Standardization of Material Declaration*, Japan Green Procurement Survey Standardization Initiative, 2nd Ed., June 2004.
10. *Material Composition Declaration of Electronic Products: Joint Industry Guide (JIG)* (Arlington, VA: Electronic Industries Association (EIA), April 2005).
11. "Guidelines for Limiting Exposure in Time-Varying Electric, Magnetic, and Electromagnetic Fields (up to 300 GHz)", International Commission on Non-Ionising Radiation Protection, (ICNIRP), *Health Physics*, 74 (1998): 494–522.
12. B. S. Collins, "RADHAZ: The Unmentionable Hazard?" *Electronics World* (October 2004).
13. *Establishing a Dialogue on Risks from Electromagnetic Fields* (Geneva: World Health Organization, 2002).
14. J. E. Moulder, K. R. Foster, L. S. Erdreich, and J. P. McNamee, "Mobile Phones, Mobile Phone Base Stations, and Cancer: A Review," *Int J Rad Biol*, vol. 81 (2005): 189–203.
15. The most current information will always be found on the web sites of the major national public health institutions, for example, <http://www.fcc.gov/oet/rfsafety/>.
16. S. Shibuya, *Basic Atlas of Radio Wave Propagation* (New York: John Wiley, 1987).
17. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications* (New York: McGraw-Hill, 2001).
18. C. Kilgus, "Shaped-Conical Radiation Pattern Performance of the Backfire Quadrifilar Helix," *IEEE Trans. AP*, vol. 23, no. 3 (May 1975): 392–397.
19. O. Leisten, J. C. Vardaxoglou, P. McEvoy, R. Seager, and A. Wingfield, "Miniaturised Dielectrically-Loaded Quadrifilar Antenna for Global Positioning System (GPS)," *Electronics Letters*, vol. 37, no. 22 (October 25, 2001): 1321–1322.
20. N. E. Lindenblad, (October 1940): U.S. Patent 2 217 911.
21. S. B. Cohn, "Optimum Design of Stepped Transmission-Line Transformers," *IEEE Trans MTT*, vol. 3, issue 3 (April 1955):16–20.
22. C. W. Davidson, *Transmission Lines for Communications* (New York: Palgrave Macmillan, 1989).
23. K. Chang, *Handbook of Microwave and Optical Components: Vol. 1, Microwave Passive and Antenna Components* (New York: John Wiley, 1997).
24. G. L. Matthai, L. Young, and E. Jones, *Microwave Filters, Impedance Matching Networks and Coupling Structures* (Norwood, MA: Artech House, 1980).
25. B. S. Collins, "The Effect of Imperfect Antenna Cross-Polar Performance on the Diversity Gain of a Dual-Polarized Receiving System," *Microwave Journal*, vol. 43, no. 4 (April 2000): 84–94.
26. *Control Interface for Antenna Line Devices*, Standard AISG v2.0, Antenna Interface Standards Group (AISG Ltd.), 2006. Available from www.aisg.org.uk.
27. C. G. Parini et al, *Semi-Smart Antenna Technology*, Ofcom Document No. 830000081/03 (London: Ofcom, 2006).

Chapter 30

Portable TV Antennas

Mitsuo Taguchi
Nagasaki University

CONTENTS

30.1	INTRODUCTION.....	30-2
30.2	DIGITAL TERRESTRIAL TELEVISION IN THE WORLD	30-2
30.3	MINIATURIZATION TECHNIQUES FOR PORTABLE TV ANTENNAS.....	30-3
30.4	RECEPTION ANTENNA MOUNTED ON VEHICLE	30-6

30.1 INTRODUCTION

Digital terrestrial television broadcasting is being introduced in many countries. This is providing a variety of services on a single channel, such as high-definition television (HDTV) and multimedia data. Obtaining portable TV reception generally is difficult compared to obtaining fixed reception at home. For fixed reception, a directional antenna with high gain and a narrow beam width, such as the Yagi-Uda dipole array antenna, is used.¹ For portable reception, an omnidirectional antenna or low-gain antenna is used. The objective of this chapter is to describe portable reception antennas suitable for digital terrestrial TV, such as small antennas with amplifier circuits, mobile phone antennas, and reception antennas mounted on cars.

30.2 DIGITAL TERRESTRIAL TELEVISION IN THE WORLD

Table 30-1 shows the frequency allocation plan in the world.² In many countries (but not in the United States or Canada), the high-band VHF from 174 MHz to 240 MHz and the UHF band from 470 MHz to 862 MHz are used for the digital terrestrial television.

TABLE 30-1 Frequency Allocation Plan and Channel Plan for Digital Terrestrial Television in the World

USA (Nov. 1988)	54 72 76 88 [Ch. 2-4] [5-6]	174 216 [Ch. 7-13]	470 608 614 698 MHz [Ch. 14-36] [Ch. 38-51]
Canada (March 2003)	54 72 76 88 [Ch. 2-4] [5-6]	174 216 [Ch. 7-13]	470 608 614 698 MHz [Ch. 14-36] [Ch. 38-51]
England (Sept. 1998)			470 550 630 806 MHz [Ch. 21-30] [Ch. 41-62]
Ireland (2006)		174 209 [Ch. 5-9]	470 862 MHz [Ch. 21-69]
France (March 2005)		174 230 [Ch. 5-12]	470 830 MHz [Ch. 21-65]
Spain (Nov. 2005)			470 862 MHz [Ch. 21-69]
Portugal (unknown)		174 202 [Ch. 5-6,8]	470 582 862 MHz [Ch. 21-69]
Germany (Nov. 2002)		174 239	470 862 MHz
Finland (Aug. 2001)		174 230 [Ch. 5-12]	470 790 MHz [Ch. 21-60]
Denmark (March 2006)		174 230 [Ch. 5-12]	470 862 MHz [Ch. 21-69]
Netherland (April 2003)		174 216 [Ch. 5-10]	470 862 MHz [Ch. 21-69]
Sweden (April 1999)		174 230 [Ch. 5-12]	470 862 MHz [Ch. 21-69]
Norway (2006-2007)		174 240	470 862 MHz [Ch. 21-69]
Italy (April 2004)		174 230	470 854 MHz
Japan (Dec. 2003)			470 710 MHz [Ch. 13-52]
Korea (Oct. 2001)		174 216 [Ch. 7-13]	470 752 MHz [Ch. 14-60]
Taiwan (July 2004)			530 602 MHz [Ch. 24-35]
HongKong (-2007)			470 806 MHz [Ch. 21-62]
Singapore (Feb. 2001)			470 614 790 MHz [Ch. 24-38] [Ch. 39-60]
Australia (Jan. 2001)		174 230 [Ch. 6-12]	520 610 750 806 MHz [Ch. 28-] [Ch. 67]

30.3 MINIATURIZATION TECHNIQUES FOR PORTABLE TV ANTENNAS

This section shows examples of small portable TV reception antennas such as the small antenna with amplifier circuit and the mobile phone antenna.

Small Antenna with Amplifier Circuit for UHF-TV

Figure 30-1 shows the structure of a coplanar waveguide (CPW)-fed dipole antenna and a loop antenna for a television receiver and DC-equivalent biasing circuits.^{3,4} The antenna element and CPW are printed on a polyimide film of thickness $45\text{ }\mu\text{m}$. The relative permittivity of film is 3.5. The characteristic impedances of CPW in two antennas are chosen as $75\text{ }\Omega$ for impedance matching to the coaxial feeder of the television receiver. The antenna element and CPW are covered by a film of thickness $50\text{ }\mu\text{m}$ for protection. The silicon transistor 2SC2585 or 2SC3604 is integrated at the feedpoint of the dipole or loop antenna, respectively.

One of the primary problems with the active antenna is that the unwanted radiation from the amplifier circuit and the bias line will perturb the antenna characteristics. To overcome this problem, the amplifier circuit is mounted on CPW. CPW also serves as a bias supply line to the amplifier circuit. The available power gain of 2SC2585 is 8 to 14 dB at frequencies from 90 to 770 MHz. The nominal value of the noise figure of this transistor is about 1.6 dB at 2 GHz. To suppress the undesired radiation from the amplifier circuit, the fixed-bias circuit with collector feedback, the simple biasing circuit, is adopted for the active dipole antenna. On the other hand, the available power gain of 2SC3604 is 24 to 18 dB at frequencies from 90 to 770 MHz. As the available power gain of 2SC3604 is larger than one of 2SC2585, the self-bias circuit is adopted to suppress oscillation and to increase stability. The nominal value of the noise figure of this transistor is about 1.6 dB at 4 GHz.

The input impedance and radiation field of dipole and loop antennas are calculated by using the computer program WIPL-D based on the Method of Moment.⁵ In the calculation, antenna elements are approximated by the wire antenna whose radii are $W_i/4$ ($i=1,2$). As the dielectric film is very thin, the existence of film is not considered.

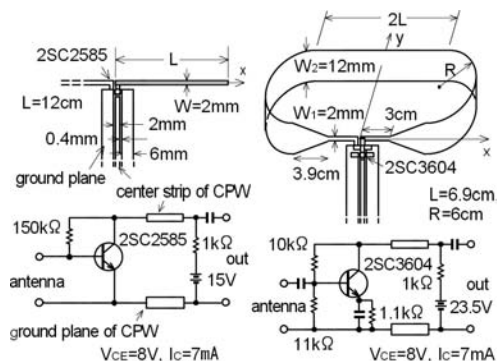


FIGURE 30-1 Dipole and loop antennas loaded with amplifier circuits

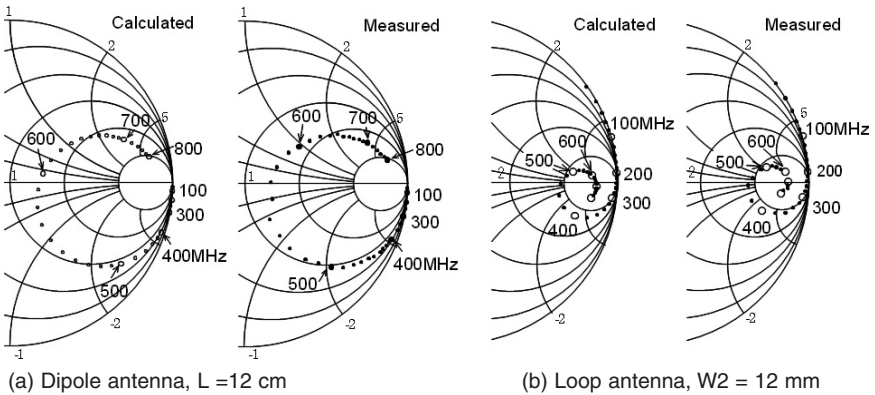


FIGURE 30-2 Input impedance of dipole and loop antennas

Figure 30-2 shows the input impedance characteristics of dipole and loop antennas. The calculated input impedances fairly agree with the measured results. Figure 30-3 shows the scattering parameters of transistors for active antennas. In the television 1–12 channel frequencies 90 to 222 MHz, the dipole antenna is not matched to the transistor circuit. Although the input reactances of dipole antennas are capacitive in the frequencies 90 to 222 MHz, those of loop antennas are inductive. This means that the conjugate impedance matching between antenna and amplifier is improved in the loop antenna. Figure 30-4 shows the calculated and measured actual gains expressed by relative values to the half-wave dipole antenna. As the impedance matching between a dipole of 24cm in length and transistor 2SC2585 is reasonable at the television 13–62 channel frequencies from 470 to 770 MHz in Japan, actual gains of more than 8 dBd are obtained. In the active

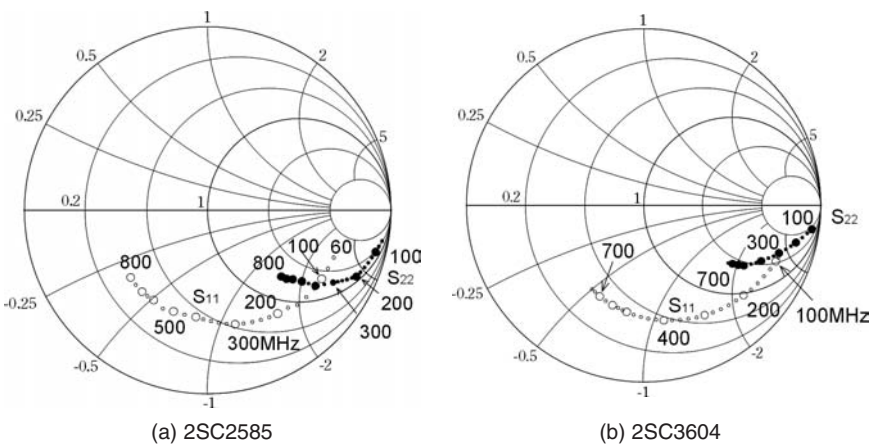


FIGURE 30-3 S parameters of silicon transistors

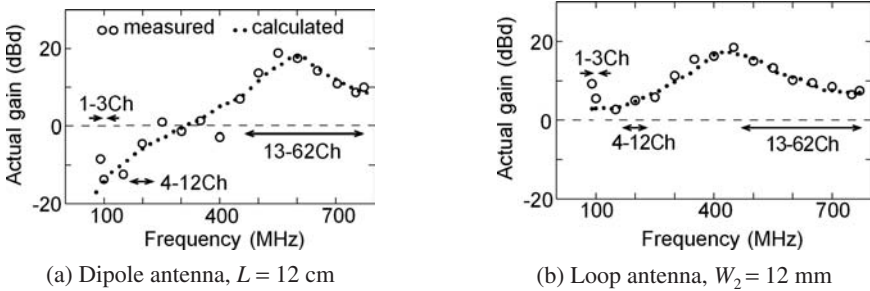


FIGURE 30-4 Actual gains of dipole and loop antennas

loop antenna, the impedance matching between loop and transistor 2SC3604 is improved in the frequencies 90 to 222 MHz. Actual gains of more than 6.5 dBd are obtained for 13–62 channels, from 470 to 770 MHz.

Portable Antenna on Mobile Phone

Digital terrestrial television broadcasting in Japan is designed so that each channel is divided into 13 segments. Among these 13 segments, 12 segments are used for HDTV transmission and 1 segment is used for mobile receivers.⁶ Maximum video resolution is 320 by 240 pixels, and maximum video bit rate is 128 kbit/s. Maximum audio bit rate is 64 kbit/s and the additional data broadcasting occupies the remaining 60 kbit/s.

Digital Multimedia Broadcasting (DMB) is a digital radio transmission system for sending multimedia (radio, TV, and datacasting) to mobile devices such as mobile phones. Currently, DMB is being put into use in a number of countries such as South Korea, Germany, France, Switzerland, China, the UK, and Indonesia.⁷

Figure 30-5 shows an example of the embedded antenna for digital terrestrial broadcasting in Japan (one segment).⁸ Its size is 38 mm by 10 mm by 2 mm. Figure 30-6 shows the antenna module, consisting of the antenna, the tuning circuit, and the low noise amplifier (LNA). Figure 30-7 shows the antenna tuning. The antenna is directly connected with the LNA, and is tuned to be synchronized with the channel selection. The antenna gain is self-controlled with the sensitivity of the tuner module.

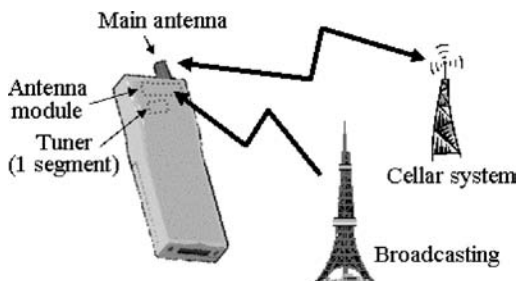


FIGURE 30-5 Antenna module for Japanese digital mobile TV

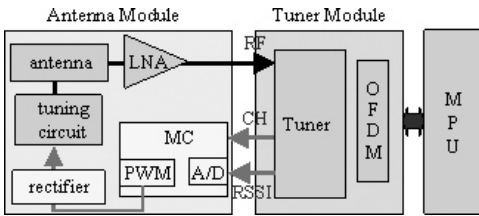


FIGURE 30-6 Block diagram of antenna module

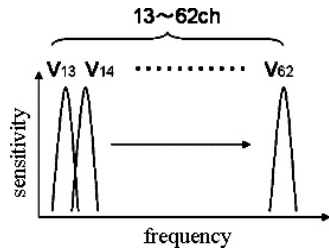


FIGURE 30-7 Antenna tuning

30.4 RECEPTION ANTENNA MOUNTED ON VEHICLE

This section shows examples of reception antennas mounted on vehicles, such as a window glass antenna, a diversity reception antenna system, and an adaptive beam-steering reception system.

Reception Antenna on Rear Window of Sedan-Type Car

Figure 30-8 shows an example of window-glass antenna printed on the rear window of a car, composed of four diversity reception antenna elements.⁹ TV antennas No. 1 and No. 2 are located close to the defogger (heater) element. TV antenna No. 4 is directly connected to a defogger element. Figure 30-9 shows the measured electric field radiation patterns in the horizontal plane. Figure 30-10 shows the average gain around all directions in the horizontal plane. The directivity is measured with a 10 dB amplifier circuit. In these figures, the angle is measured from the front direction.

Diversity Reception Antenna System

Figure 30-11 illustrates the diversity reception antenna system for TV.¹⁰⁻¹² The antenna elements are printed symmetrically on the left and right quarter windows of the car. The antenna consists of two parts: the slanted element consisting of three thin lines located

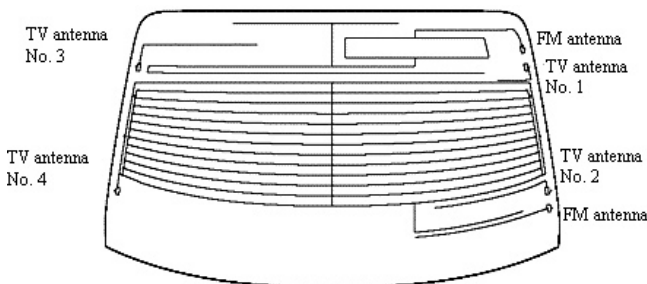


FIGURE 30-8 Window glass antenna printed on the rear window of sedan-type car

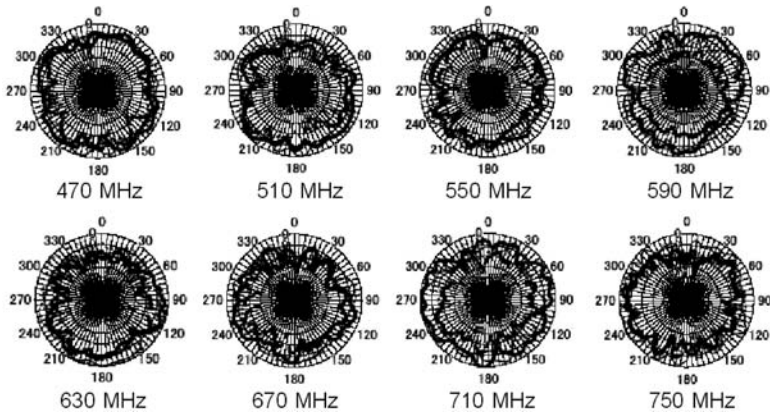


FIGURE 30-9 Measured electric field radiation patterns in the horizontal plane

close to the rear pillars, and a horizontal element printed at the center of the quarter glass. A switching diversity system was adopted because the synthetic diversity system was not sufficiently effective at suppressing multipath distortion. The switching diversity system can effectively follow changes in field strength by selecting appropriate antennas for better reception. The antenna switching is performed during every vertical blanking period in the TV video signals.

The diversity reception system for the terrestrial TV has been on the market since the middle 1980s.^{11,12} Two V-shaped pole elements were first used for TV reception. A new antenna system consisting of four antenna elements printed on the rear quarter window glass was developed in 1986. Subsequently, an adaptive array antenna for terrestrial digital TV broadcasting was developed, and you will soon be able to enjoy TV programs in high-quality digital while riding in an automobile.¹³

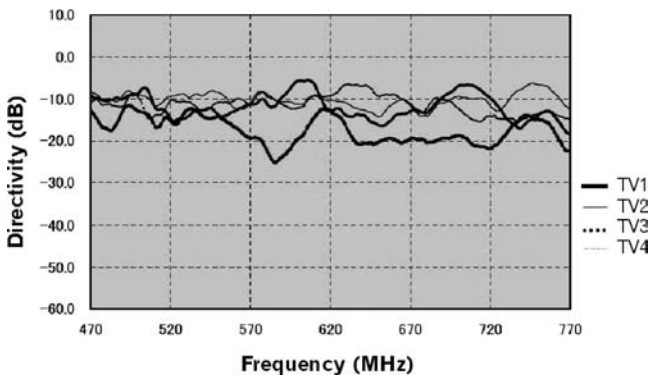


FIGURE 30-10 Average gain around all directions in the horizontal plane

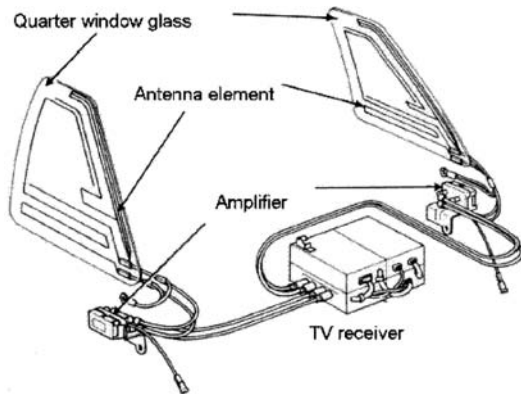


FIGURE 30-11 Diversity antenna system for TV

On-glass Mobile Antennas for Digital Terrestrial Television

Figure 30-12 shows the geometry of the horizontally polarized on-glass mobile antenna.¹⁴ The antenna has been developed by modifying a batwing antenna.¹⁵ Batwing antennas have a broad bandwidth but are too large to be installed in vehicles, so development has focused on size reduction. The proposed antenna is normal to ground and fed by a coaxial cable. The antenna height, H , was reduced by bending the wires in that direction without narrowing the bandwidth. The height of the new antenna is approximately half of the original antenna height. The prototype antenna for field experimentation using a passenger van is shown in Figure 30-12b. Two antenna elements and a mesh ground plane were printed on the inside of the rear glass window of the passenger van. The dielectric constant of the glass was 6.7, and the loss tangent was 0.008. The thickness was about 5 mm, which is very thin compared with the wavelength of the UHF band. Experiments confirmed that the glass had a very small effect on antenna characteristics. The patterns were printed using silver paste, and the mesh ground plane was connected electromagnetically to the vehicle body upon installation of the window. The conductivity of the silver paste was about 2×10^6 S/m, which is lower than that of copper (5.813×10^7 S/m). Antenna gain loss was confirmed to be less than 0.3 dB within the frequency band by comparison with the measured results of an antenna composed of copper.

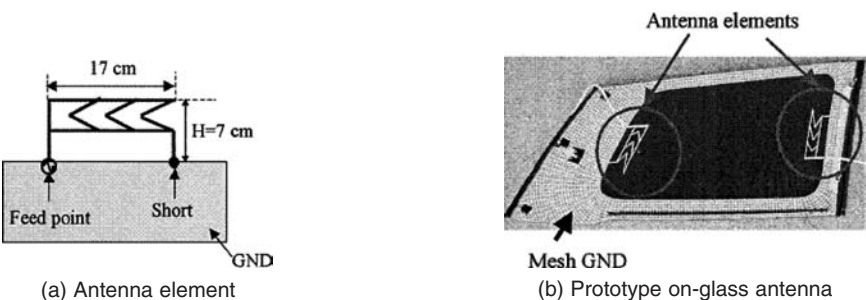


FIGURE 30-12 Geometry of on-glass antenna

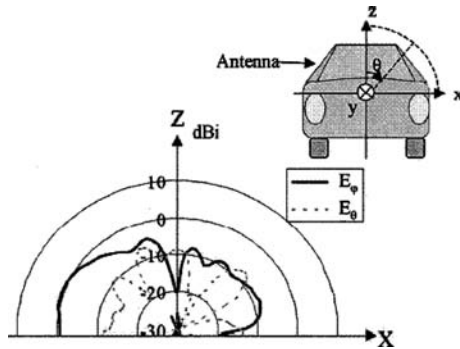


FIGURE 30-13 Measured radiation pattern (vertical plane, 485 MHz)

Figure 30-13 shows the radiation patterns of the antenna installed on the left side of the passenger van at 485 MHz. As expected, the elevation pattern peak of E_ϕ , the co-polarization of DTV, was in the horizontal plane. Figure 30-14 shows the radiation patterns in the horizontal plane at 485 MHz. The antenna element printed inside the left side glass radiates strongly to the left side of the vehicle, and the element inside the right side glass radiates strongly to the right side. A peak gain of 4.5 dBi was achieved by each antenna element, further showing the suitability of this type of antenna for use in a diversity system.

A diversity system can achieve a high C/N ratio in built-up urban areas where multipath fading is strong. Maximum Ratio Combining (MRC) was selected as the signal processing method.¹⁶ The radiation pattern of MRC was estimated by combining the measured radiation patterns of four antennas installed on the passenger van, as shown in Figure 30-15. An average gain of 4.6 dBi was achieved on the horizontal polarization.

Figure 30-16 shows the average gain of the pattern. The solid line represents the gain of the proposed antenna and the dotted line represents a comparison using the average gain of four commercial pole antennas installed on the roof. These pole antennas were leaned at an angle of 45 degrees to the roof. The average gain of four crossed-dipole antennas installed on the corner of the roof was also measured at 485 MHz.¹⁷ From 470 MHz to 710 MHz, the average gain of the proposed antenna was greater than 3.1 dBi. This was 4.7 dB higher than that of the commercial pole antennas, which is the minimum value within the bandwidth. It was also 1.6 dB higher than the average gain of crossed-dipole antennas at 485 MHz.

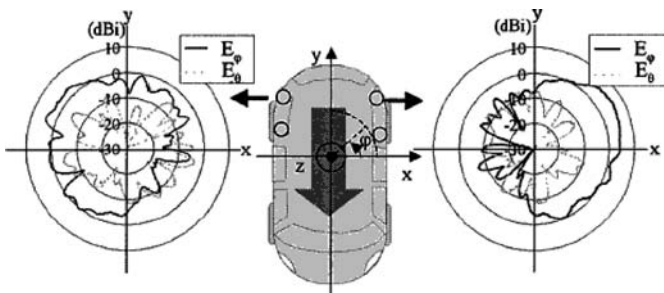


FIGURE 30-14 Measured radiation pattern (horizontal plane, 485 MHz, element)

Modified H-shaped Antenna

Figure 30-17 shows a modified H-shaped antenna for automotive digital terrestrial reception.¹⁸ Two wires $a-b-c-d-e$ and $f-g-h-i-j$ are symmetrically placed around the center m of the antenna and connected by the wire $c-m-h$. The parts $a-b-c$ and $f-g-h$ of the two wires are longer than the parts $c-d-e$ and $h-i-j$. The antenna is placed in the xy plane. A control system for combining received signals is shown in Figure 30-17b. The received signals of the four antennas are down-converted, weighted, and combined. The weight vector for each signal is controlled, based on the MRC method. The four antennas mounted at the top of the front and rear windows are also depicted. The antenna plane, which includes the antenna element, is inclined from the horizontal plane in an actual car.

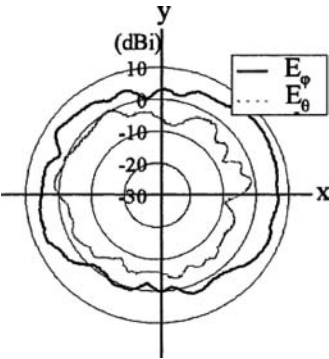


FIGURE 30-15 Radiation pattern (four-element array)

A design concept of the coverage in the xy plane is summarized in Figure 30-18. Mechanism of the antenna is described with Figure 30-17a and Figure 30-18. The antenna has three resonant modes. Each resonant mode is excited when a part of the antenna becomes a half-wavelength. The frequency band for digital terrestrial services is divided into three bands. The wire $a-b-c-m-h-g-f$ is resonated at the low frequency, which is series resonance. Parallel resonance occurs at the middle frequency. The two wires $a-b-c-d-e$ and $f-g-h-i-j$ are resonated. Series resonance occurs again at the high frequency, in which the wire $e-d-c-m-h-i-j$ is resonated. The radiation pattern of a single element is rotated clockwise with increasing frequency. The minimum level of relative amplitude increases when the maximum radiation is directed toward the x -axis direction. The four antennas are symmetrically placed around the x -axis and y -axis directions in Table 30-2, which is assumed to be equivalent to installation at the top of the front and rear windows. The symmetrical arrangement allows coverage of 360 degrees.

The wires shown in Figure 30-17a were etched on a typical FR4 substrate having dielectric constant of 4.6 and thickness of 0.8 mm. The width of the lines was 1 mm. The lengths L_L , L_S , and L_C were experimentally adjusted to 102, 67, and 30 mm, respectively, with constant value of $H = 60$ mm, since the value of H was specified for car installation. The prototype antenna has LC balun and a parallel feedline between the LC balun and the feedpoint m .

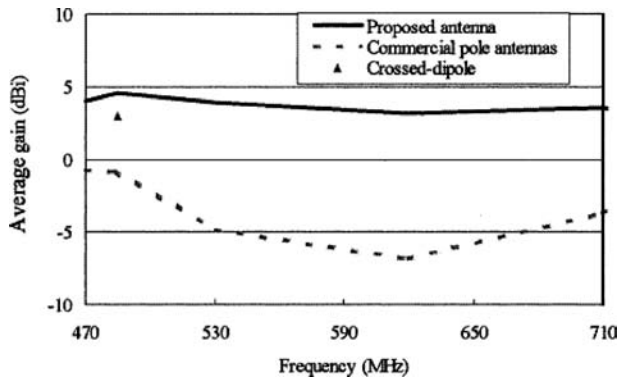


FIGURE 30-16 Average gain of the adaptive array pattern (E_ϕ)

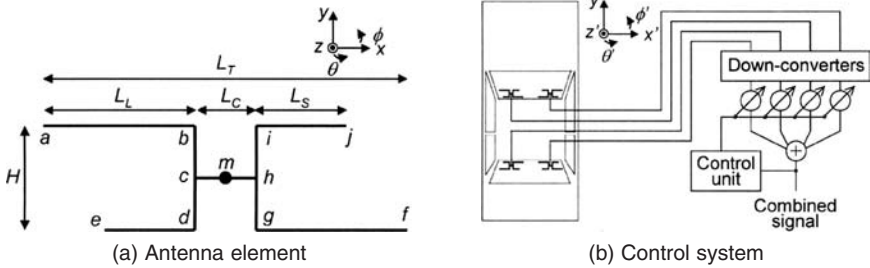


FIGURE 30-17 Modified H-shaped antenna and control system for combining received signals for automotive digital terrestrial reception

The LC balun transforms impedance from $110\ \Omega$ to $50\ \Omega$ as well as balance-unbalance mode. The LC balun consists of a high-pass filter (HPF), a low-pass filter (LPF), and a T-junction. Both filters have the fifth-order Butterworth function. A 3 dB cutoff frequency is set at a low frequency of 272 MHz for the HPF and a higher frequency of 1228 MHz for the LPF of the frequency band from 470 to 710 MHz so that both filters may give flat amplitude and out of phase. Insertion loss of the LC balun was measured to be 0.3 dB across the frequency band. The parallel feedline has a length of 40 mm. The width of each line and the distance between the centers of the two lines were chosen to be 0.8 and 1.1 mm, so that characteristic impedance would be $110\ \Omega$.

Figure 30-19 shows the measured radiation patterns at the three frequencies. The figure-8 radiation patterns were rotated clockwise with increasing frequency as the calculation results predicted. VSWR was measured to be less than 3 from 470 to 710 MHz.

Four prototype antennas were mounted at the top of the front and rear windows of the car, as shown in Figure 30-17b. The distance between the edges of the metal roof and the nearest side of the prototype antennas was set at 20 mm. The spacings of antenna elements were, respectively, 600 mm in the x' -axis direction and 1600 mm in the y' -axis direction. Figure 30-20a–d show typical measured radiation patterns of the four prototype antennas in the ϕ' plane at 530 MHz. Each prototype antenna had gain to the x' -axis direction as well as to the y' -axis direction because the four antennas in free space had inclined figure-8 radiation patterns. The peak plot combined with the radiation patterns in Figure 30-20a–d is shown in Figure 30-20e. A near omnidirectional pattern was achieved in the ϕ' plane. In terms of vertical plane, gain of the peak plot decreased with increasing the angle θ' ,

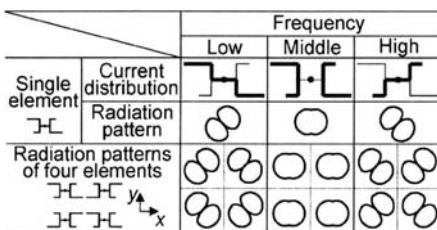


FIGURE 30-18 Concept covering 360 degrees in the xy plane using the four antennas across the frequency band for DTV

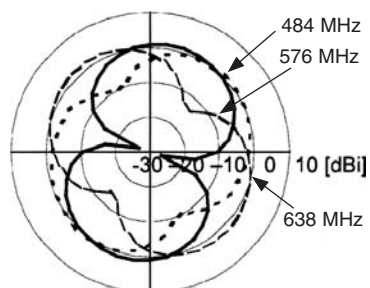


FIGURE 30-19 Measured radiation patterns of a prototype antenna in the xy plane

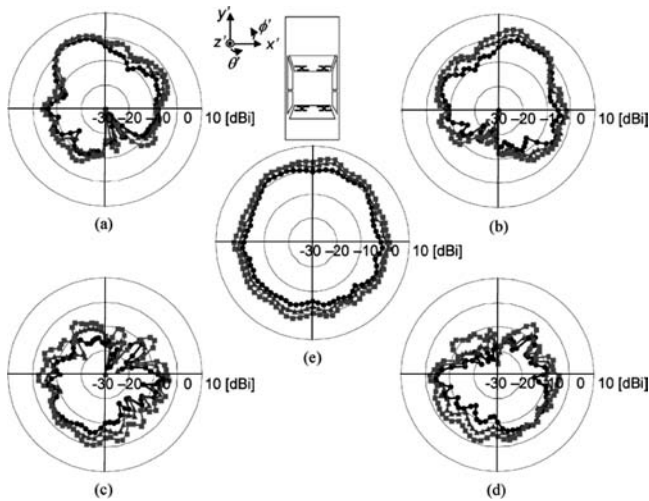


FIGURE 30-20 Measured radiation patterns in ϕ' plane at 530 MHz. Four prototype antennas were mounted at the top of the front and rear windows of a car. Peak plot of the combined pattern based on the MRC method is also presented.

because radiation patterns of the mounted prototype antennas were affected by the effect of the car body. The values of the minimum gain of the peak plot were -2.3 dBi for $\theta' = 70$ degrees, -3.7 dBi for $\theta' = 80$ degrees, and -6 dBi for $\theta' = 90$ degrees.

Adaptive Beam Steering Reception System

To improve the reception quality of television on mobile terminals, the adaptive beam steering system has been studied in Japan.¹⁹

Figure 30-21 shows the adaptive beam steering reception system. This system adopts the diversity technique, combining received signals prior to FFT processing on the OFDM demodulator. The experiment was conducted in the Nagoya suburbs using a van with four antennas on the windows, and was focused on traveling on highways. From the experiment, it was shown that the mobile reception area using the adaptive beam steering system is almost the same as the fixed reception area.¹⁹

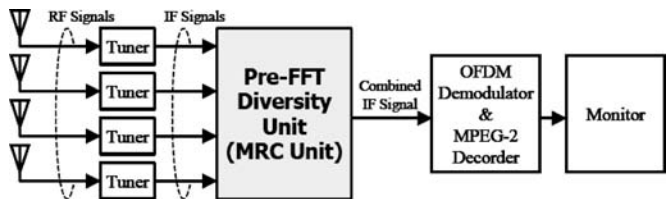


FIGURE 30-21 Adaptive beam steering reception system

REFERENCES

1. E. B. Joy, *Antenna Engineering Handbook*, Chap. 29, R. C. Johnson (ed.) (New York: McGraw-Hill, 1993).
2. Federal Communications Commission, "FCC Online Table of Frequency Allocations," <http://www.fcc.gov/oet/spectrum/table/fcctable.pdf> (revised November 29, 2006).
3. M. Taguchi, T. Fujimoto, and K. Tanaka, "CPW Fed Active Dipole Antennas for Television Receivers," *Electronics Letters*, vol. 30, no. 22 (October 1994): 1815–1816.
4. M. Taguchi and T. Fujimoto, "Actual Gain of CPW-Fed Active Integrated Antennas for Television Receiver," *IEICE Trans. Commun.*, vol. E81-B, no. 7 (July 1998): 1542–1547.
5. WIPL-D, <http://www.wipl-d.com/>.
6. Association of Radio Industries and Businesses, *Transmission System for Digital Terrestrial Television Broadcasting*, ARIB STD-B31, ver. 1.6 (November 2005; in Japanese).
7. Wikipedia, s.v. "Digital Multimedia Broadcasting," http://en.wikipedia.org/wiki/Digital_Multimedia_Broadcasting.
8. K. Takei and T. Sugiyama, "Antenna Module Self-Controlling Gain with Sensitivity Feedback," Proc. of Colloquium on RF for DVB-H/DMB Mobile Broadcast: Handset and Infrastructure Challenges, London, June 2006.
9. Technical data described in this section is provided by courtesy of Central Glass Co., Ltd.
10. K. Nishikawa, "Land Vehicle Antennas," *IEICE Trans. Commun.*, vol. E86-B, no. 3 (March 2003): 993–1004.
11. H. Toriyama, J. Ohe, H. Kondo, and H. Yotsuya, "Development of Printed-on Glass TV Antenna System for Car," *Proc. IEEE Vehicular Tech. Conf.* (1987): 334–342.
12. J. Ohe, "Development of Printed-on Quarter Glass Antenna System for Cars," *ITEJ Technical Report*, vol. 13, no. 26, RE'89-15 (May 1989): 19–24.
13. J. Imai, M. Fujimoto, N. Itoh, T. Shibata, N. Suzuki, and K. Mizutani, "Mobile Reception Experiments of Digital Terrestrial Broadcasting—A Study on Beam Control for Improvement of Reception Quality," Proc. Commun. Conf. IEICE 2001, B-5-113, September 2001.
14. S. Matsuzawa, H. Sato, and K. Nishikawa, "Study of On-glass Mobile Antennas for Digital Terrestrial Television," *IEICE Trans. Commun.*, vol. E88-B, no. 7 (July 2005): 3094–3096.
15. R.C. Johnson, (ed.), *Antenna Engineering Handbook* (New York: McGraw-Hill, 1984).
16. M. Fujimoto, K. Nishikawa, T. Shibata, N. Kikuma, and N. Inagasaki, "A Nobel Adaptive Array Using Frequency Characteristics of Multi-carrier Signals," *IEICE Trans. Commun.*, vol. E83-B, no. 2 (Feb. 2000): 371–379.
17. J. Imai, M. Fujimoto, T. Shibata, N. Suzuki, N. Itho, and K. Mizutani, "Experimental Results of Diversity Reception for Terrestrial Digital Broadcasting," *IEICE Trans. Commun.*, vol. E85-B, no. 11 (November 2002): 2527–2530.
18. H. Iizuka et al., "Modified H-Shaped Antenna for Automotive Digital Terrestrial Reception," *IEEE Trans. Antennas and Propagat.*, vol. 53, no. 8 (August 2005): 2542–2548.
19. K. Sanda, T. Shibata, N. Itoh, K. Ito, N. Suzuki, and J. Imai, "Adaptive Beam Steering Reception System for ISDB-T Based on Pre-FFT Diversity Technique," Proc. IEEE Int. Conf. on Consumer Electronics (ICCE), 11.2-2, Las Vegas, January 2006.

Chapter 31

Reconfigurable
Antennas

Jennifer T. Bernhard

University of Illinois at Urbana-Champaign

CONTENTS

31.1	INTRODUCTION	31-2
31.2	RECONFIGURABLE APERTURES	31-2
31.3	RECONFIGURABLE TRAVELING WAVE ANTENNAS	31-5
31.4	RECONFIGURABLE ARRAYS	31-8
31.5	RECONFIGURABLE MICROSTRIP ANTENNAS	31-10
31.6	RECONFIGURABLE SLOT ANTENNAS	31-13
31.7	RECONFIGURABLE MONOPOLE/DIPOLE ANTENNAS	31-16
31.8	CONCLUDING REMARKS	31-19

31.1 INTRODUCTION

Wireless systems are evolving toward multifunctionality. This multifunctionality provides users with options of connecting to different kinds of wireless services for different purposes at different times, whether in a city or on a battlefield. To reach their full potential, these systems require not only agile electronics and adaptive signal processing, but also reconfigurable antennas. Reconfigurable antennas can be used simply to reduce the number of antennas necessary for intended system function, but they can also be designed to serve much more complex roles. Some of these roles include use as programmable control elements with feedback to increase throughput, reduce errors and noise, and improve security, and use as reconfigurable hardware to extend the lifetime of the entire system. Examples of emerging applications include software defined radio, cognitive radio, multiple-input multiple-output (MIMO) systems, multifunction consumer wireless devices, and high performance phased arrays.

As one examines the development of reconfigurable antennas, and, in fact, antennas in general, one can see that each antenna is driven by specific applications or functionality. Indeed, there is no one ultimate application that would result in a super antenna that would be able to meet the specifications of every system. In this, antenna engineers are truly fortunate—practical issues, such as cost, weight, size, and complexity, make each new system or application a new challenge. Antenna reconfigurability can help to meet these challenges in new and exciting ways. This chapter describes the work of talented engineers and researchers in their efforts to develop reconfigurable antennas, from large apertures for communication and sensing satellites to single individual elements on small portable wireless devices. These examples, as well as others, are the first steps in establishing a strong foundation of antennas for fully multifunctional systems.

A Brief Definition of Antenna Reconfigurability

A reconfigurable antenna is one that has selectable or tunable fundamental characteristics, including operating frequency, impedance bandwidth, radiation pattern, and polarization. Typically, the goal is to change one or more of these characteristics independently of the others. The methods in which these characteristics are changed are electrical, mechanical, or electromechanical in nature, since fundamentally, antennas are electromechanical structures. However, given the wide range of intended applications and design specifications, engineers have developed an equally wide and varied collection of reconfiguration mechanisms.

In each of the following sections, theoretical foundations and examples of different kinds of reconfigurable antennas are presented. While there are many more reconfigurable antennas than can be discussed here, the selected subset of antennas is intended to provide a broad sampling of the rich history and future potential of these tools for communication, sensing, and security.

31.2 RECONFIGURABLE APERTURES

In concept, reconfigurable apertures represent the ideal functionality of reconfigurable antennas in that they allow the complete specification of radiating currents that can produce any desired frequency and radiation behavior. The challenge lies in developing real structures and control mechanisms that can support this functionality.

Motivation for Reconfigurable Apertures

As mentioned previously, each antenna is typically designed to deliver functionality for a specific application. In the case of aperture antennas, they are most often implemented for long distance, high gain applications intended for more expensive and specialized systems, such as satellites, long-haul land-based communications, and radar receivers. However, when the antenna is large in extent and is deployed in a hostile or remote location, reconfigurability plays a different role. In this case, it could lengthen the useful lifetime of the system by allowing for reprogramming/retasking, without the need for retooling or for a labor- and cost-intensive maintenance visit. The apertures discussed here implement a variety of approaches to enable this kind of reconfigurability.

Examples of Reconfigurable Apertures

Reflector antennas are perfect candidates for radiation pattern reconfiguration independent of frequency, since the reflective surface is physically removed and isolated from the primary feed. Mechanical deformations of the reflector (or subreflector) surface result in changes in the surface currents on the reflector, which in turn alter the aperture fields, and finally, the radiation pattern of the antenna. In one of the earlier examples of antenna reconfigurability using mechanical means, Clarricoats and his collaborators demonstrated radiation reconfiguration with a reflector antenna by changing the structure of a mesh reflector. Initially, manual adjustments in each meshed region of interest served to change the overall reflector contour, resulting in changes in beam shape and direction.¹ Later, the implementation of computer-controlled stepper motors, used for pulling lines attached to specific points on the reflector mesh, allowed for automatic pattern reconfiguration via electromechanical actuation.²

Recently, another group developed and expanded the functionality of a similar system for satellite applications by changing the shape of a subreflector, with the shape of the main reflector remaining fixed.³ A thin, flexible, conductive material composes the subreflector. The subreflector's shape is changed using piezoelectric actuators attached to its back surface. As the actuators deform the surface of the subreflector, the electromagnetic field illuminating the main reflector changes, which results in a desired radiation pattern. Figure 31-1 shows a subreflector system schematic.³ Enabling changes in the subreflector, rather than the main reflector, simplifies the antenna system fabrication and allows designers to implement main reflectors of varying geometries and deployments. This is especially important for reducing probabilities of failure during deployment or operation for space applications. One important goal of the antenna in this case³ is to achieve the desired reconfigurability with the minimum number of actuators. During the antenna system design, the positions of the actuators are determined using an iterative finite element algorithm. This algorithm determines where successive actuators should be placed in order to minimize the error between the desired and actual subreflector shape.³ In providing the ability to change the antenna's radiation pattern, this design approach can be used to effectively lengthen the useful lifetime of communication and relay satellites.

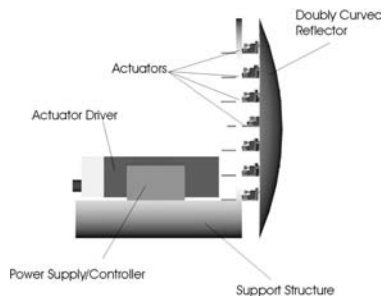


FIGURE 31-1 Reconfigurable subreflector system, including piezoelectric actuators and driver to deflect the surface of a Cassegrain subreflector (after G. Washington et al³ © IEEE 2002)

Moreover, it can dramatically reduce cost by allowing antennas to be retasked via programming from the ground rather than requiring completely new satellites to enable new functions and all of the cost that this entails.

The next two examples of reconfigurable antennas may stretch the definition of an aperture in the traditional sense, but one can certainly consider them to be blank palettes for establishing desired current distributions that result in desired antenna behavior. In essence, these apertures are intended to create whatever antenna is necessary to achieve specific frequency, bandwidth, radiation, and polarization characteristics. The first is an antenna that uses switches (which can be based in solid-state, MEMS, or other technologies) for connecting small non-resonant conductive pads that create a large antenna aperture.⁴ The number of switches required for such an aperture can easily number into the thousands. Figure 31-2 shows a conceptual drawing of the aperture.⁴ The desired antenna configuration is found using a genetic search algorithm, which determines the needed combination of switch states across the aperture for achieving a specified performance goal.⁴ For each pass through the genetic algorithm, a finite-difference time-domain electromagnetic simulation is used to evaluate candidate designs. In general, narrowband configurations achieve more gain than wideband configurations, since wideband configurations implement fewer effective radiators in the aperture, which has a fixed physical size. While this aperture approach promises tremendous flexibility in terms of both frequency and radiation performance, it does have some inherent drawbacks. High frequency operation may be degraded as a result of the pad density and switch capacitance.⁴ Additionally, for designs at lower frequencies, the fixed physical size of the aperture may not be able to support high gain operation. Moreover, while the large number of switches between pads provides for graceful degradation of operation with switch failure, the complexity of the required bias networks to allow the switches to be individually addressable may make the final product too costly and bulky for its intended application.

The second related antenna aperture concept uses semiconductor plasmas to form antenna structures.⁵ Developed by researchers at the Sarnoff Corporation, this pixel approach to a reconfigurable aperture relies on high conductivity plasma islands that are formed and

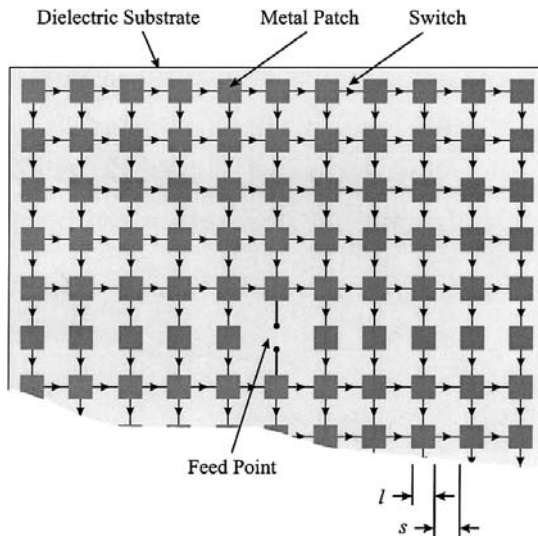


FIGURE 31-2 Conceptual drawing of a reconfigurable aperture antenna based on switched links between small metallic patches (after L. N. Pringle et al⁴ © IEEE 2004)

controlled by DC currents that have been injected into high-resistivity silicon-based diode structures. While the fundamental concept—that of creation of conducting patches in an aperture—is identical to the last antenna discussed, this antenna aperture differs significantly in the technology necessary for achieving the desired current distributions. Figure 31-3 contains a detailed description of the structure, including the plasma injection driver configuration.⁵ While elegant in its conception, the embodiment is fundamentally just as complex as the previous example that uses many discrete switches. This means that many practical and technical issues exist with this approach, including the challenge of creating the necessary carrier densities in the silicon. However, it possesses one compelling feature that other reconfigurable apertures based on conductors do not have: when turned off, the aperture possesses an extremely low radar cross-section. Similar to the previous example,³ this approach to antenna reconfigurability may be more appropriate for creating intermediate apertures that are illuminated with more conventional sources. In this case, the plasma aperture could be used to reconfigure holograms for holographic antennas in which a basic source, such as a patch or horn antenna, is used to illuminate a hologram. The aperture field for the desired pattern is then produced through interaction with the plasma aperture, enabling beamsteering and beamforming without the need for expensive phase shifters. The plasma aperture could then be designed using iterative or other means and reconfigured to deliver performance comparable to traditional phased arrays.⁵⁻⁶

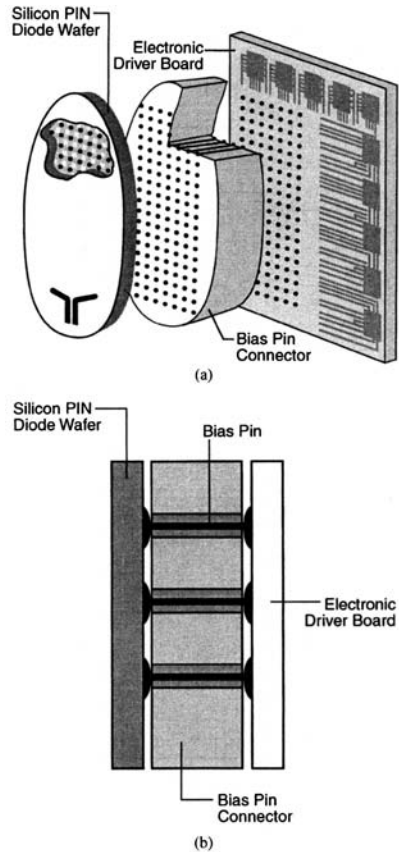


FIGURE 31-3 Depiction of a controllable plasma grid structure for a reconfigurable aperture: (a) The grid structure is fabricated on top of the silicon wafer. (b) A cross-section of the plasma grid structure shows the layer interconnection (after A. Fathy et al⁵ © IEEE 2003).

31.3 RECONFIGURABLE TRAVELING WAVE ANTENNAS

Traveling wave antennas are well known for delivering frequency-scanned radiation patterns. Introduction of reconfigurability into these structures adds another layer of functionality that can reduce some of the system costs inherent in previous designs.

Motivation for Reconfigurable Traveling Wave Antennas

Traveling wave antennas have been used with great success for delivering broadband beam scanning performance through frequency sweeps, without the need for phase shifters.

However, the necessity to change operating frequency to steer beams in a particular direction places burdens on both ends of a communication link. The reconfigurable traveling wave antennas discussed here have made real gains in delivering frequency-independent beam scanning, without compromising the effective operating bandwidth of the original base antenna.

Examples of Reconfigurable Traveling Wave Antennas

One recently-developed reconfigurable traveling wave antenna is a dielectric waveguide antenna, which relies on the mechanical perturbation of propagation constants at millimeter-wave frequencies.⁷ The widths of the gratings on a thin, moveable film placed over a dielectric image line are designed to gradually perturb the propagation constant along the line, as shown in Figure 31-4.⁷ Viewed in light of the antenna presented in the previous section, this film serves as an effective hologram for the desired pattern. Physical shifts of the grating film change the apparent grating spacings and result in a scanned beam at a single frequency. Scan angles of up to 53 degrees have been demonstrated with this design at 35 GHz, with lower scan angles achievable over an operating band between 30 and 40 GHz.⁷ An extension of this concept delivers scanned, dual beam performance over a similar frequency range.⁸ Employing the concept of the holographic antenna,⁵⁻⁶ its functionality could be expanded further by enabling reconfigurability of the film itself.

Another interesting reconfigurable traveling wave antenna, first developed in 1960, is the electromechanically scannable trough waveguide antenna.⁹ Based on an asymmetrical trough waveguide intended to feed a parabolic reflector,¹⁰ two different mechanisms were investigated to change the phase velocity in the guide, resulting in scanned radiation patterns at a single frequency. The first used the rotation of artificial dielectric structures along a longitudinal axis within the trough waveguide, and the second used a mechanical variation of the height of periodic structures located on the top of the center fin of the trough waveguide.⁹ While elegant in concept, the authors acknowledged that both approaches presented major mechanical problems. They were solved using cam and gear solutions, which proved to be somewhat cumbersome.

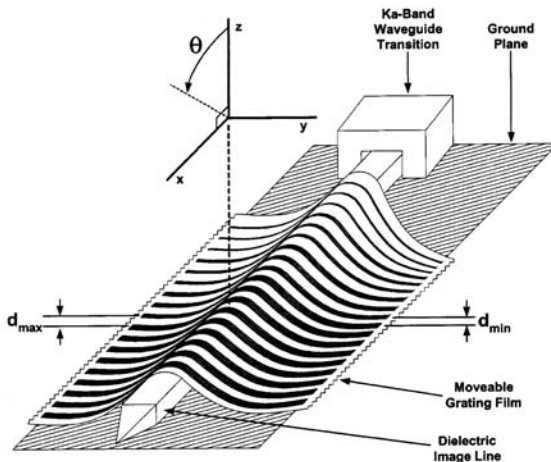


FIGURE 31-4 A traveling wave antenna based on moveable grating fed by a dielectric image line, capable of beam scanning at millimeter-wave frequencies (after K. Chang et al⁷ © IEEE 2002)

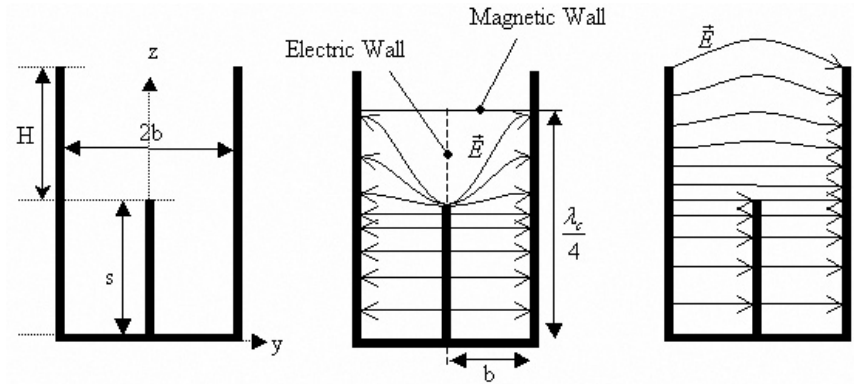


FIGURE 31-5 Trough waveguide dimensions (left), electric field lines of the dominant TE mode with equivalent structure for design and field distribution (middle), electric field lines of the TEM mode (right) (after G. H. Huff and J. T. Bernhard¹¹ © 2005)

Huff and Bernhard¹¹ begin with the concept of the electromechanically scanned trough waveguide⁹ and then propose the application of more modern perturbations and streamlined actuation mechanisms. This allows them to achieve beamsteering at a single frequency while preserving the wide bandwidth characteristics of the antenna. The fundamental structure remains the trough waveguide, however, as shown in Figure 31-5.¹¹ In this case, beamsteering is achieved using the micro-machined cantilever structures fabricated within the trough waveguide (shown in Figure 31-6),¹¹ to create perturbations suitable for electrostatic or magnetostatic actuation. Simulations with actuation of individual cantilevers or groups of cantilevers demonstrate beam scanning at single frequencies through broadside for this traveling wave antenna.¹¹

Other leaky wave antennas, which use tunable-impedance, ground planes to enable beam steering, have been developed.¹² Based on the same fundamental concept of controlling wave velocity changes as in the previous example, the structure¹² relies on a tunable-impedance surface equipped with hundreds of diodes to affect beam changes.

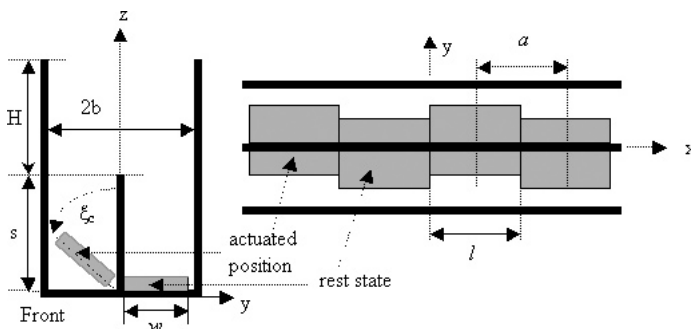


FIGURE 31-6 Trough waveguide with cantilever-type perturbations, showing staggered positions of actuated and rest states of the cantilevers (after G. H. Huff and J. T. Bernhard¹¹ © 2005)

31.4 RECONFIGURABLE ARRAYS

The signals used with phased arrays of antennas are already reconfigurable in their magnitude and phase. The addition of reconfigurability at the level of the individual array element promises to provide new levels of system performance.

Motivation for Reconfigurable Arrays

Of course, the one antenna topology that is already well known for its flexibility is the antenna array. This flexibility is traditionally delivered through control of the signal magnitude and phase of identical antenna elements, arranged either periodically or aperiodically, in a plane or conformally on a three dimensional surface. Historically used in defense-oriented applications, arrays are now being used in a much broader set of scenarios to provide additional functionality and security. The reconfigurable arrays described here expand typical array functionality in frequency and radiation performance, providing additional degrees of freedom for system designers and ultimately, system users.

Examples of Reconfigurable Arrays

One persistent issue with antenna arrays is their relatively limited frequency bandwidth. One way to improve the bandwidth of such arrays is to make the individual elements frequency reconfigurable. However, in arrays that have a ground plane, the presence of the ground plane at a fixed distance from the plane of the antennas does not enable the complete frequency reconfigurability present in the elements. In one case,¹³ a novel layered frequency-selective surface was implemented to support the performance of broadband reconfigurable antenna arrays. A diagram of the system concept is provided in Figure 13-7.¹³ With careful design, the frequency selective surfaces appear as magnetic ground planes and deliver nearly constant phase over a broad frequency band. This can enhance array gain, as shown in Figure 31-8,¹³ which would otherwise be limited by a fixed ground plane spacing. The same authors have applied this methodology to a wideband reconfigurable slot aperture.¹⁴

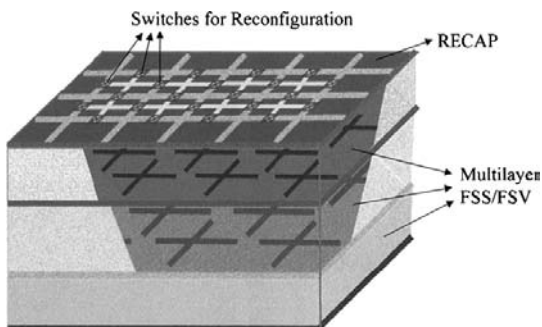


FIGURE 31-7 Depiction of a reconfigurable aperture (RECAP) array over a multilayer frequency selective surface (FSS)/frequency selective volume (FSV) structure (after Y. E. Erdemli et al¹³ © IEEE 2002)

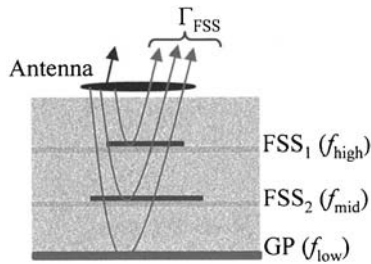


FIGURE 31-8 Diagram illustrating the coherence of direct antenna radiation and reflections from the substrate that helps to provide broadband operation of reconfigurable arrays (after Y. E. Erdemli et al.¹³ © IEEE 2002)

Using another approach, pattern reconfigurable antennas were studied in small arrays.¹⁵⁻¹⁸ The individual element in the array is a reconfigurable spiral microstrip antenna, also called a reconfigurable magnetic line source antenna. This antenna, through the inclusion of solid-state or RF MEMS switches, delivers broadside and endfire radiation patterns.¹⁵⁻¹⁸ When arrayed, these elements can extend the scan capabilities over that achievable with a traditional planar phased array.¹⁵⁻¹⁸ A photograph of a 4×4 array of these elements is shown in Figure 31-9.¹⁸ Measured radiation patterns for the array, with all of the elements in broadside mode with uniform amplitude and phasing, provide a typical broadside pattern. With all of the elements in endfire mode with uniform amplitudes and 180 degree progressive phasing, the array produces a dual-beam endfire pattern.¹⁸ In this case, the patterns of the individual elements are used to reinforce the array factor created in the usual way with element amplitude and phasing. This example provides only a small sample of the array capabilities that can be achieved using individually reconfigurable antenna elements in arrays.

Reconfigurable arrays with electrically tuned or switched parasitic elements can provide a great deal of pattern variability while maintaining impedance match and bandwidth. Fundamentally, tuning of antenna radiation patterns in this manner relies on the mutual coupling between closely spaced driven and parasitic elements, resulting in effective array behavior from a single feed point. Therefore, changes in radiation patterns are achieved through changes in the coupling between the elements, which, in turn, change the effective source currents on both the driven and parasitic elements.

One of the fundamental examples of this antenna was proposed by Harrington in 1978.¹⁹ Shown in Figure 31-10,¹⁹ the single dipole element in the center of the array is driven, while tunable reactances on the surrounding parasitic dipoles couple to the source antenna and produce a directive beam in a desired direction. A waveguide-based reconfigurable array

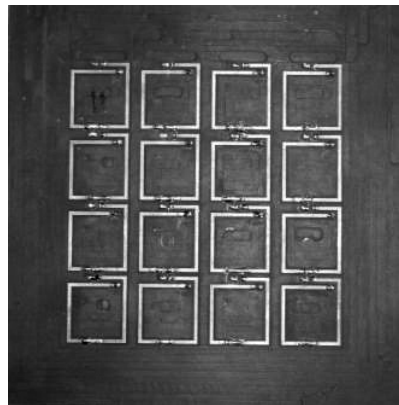


FIGURE 31-9 Photograph of a 4×4 array of pattern reconfigurable elements that can be configured to deliver broadside, endfire, and scanned radiation patterns (after G. H. Huff¹⁸ © 2006)

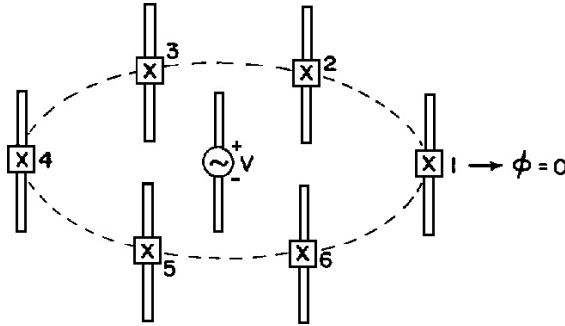


FIGURE 31-10 A seven-element circular array of reactively loaded parasitic dipoles for reconfigurable beam steering and beamforming (after R. F. Harrington¹⁹ © IEEE 1978)

was also proposed for the same concept.²⁰ Since then, a number of reconfigurable parasitic arrays have been proposed and studied, using both switched and reactively loaded elements, which provide a wide variety of functionality.^{21–24}

31.5 RECONFIGURABLE MICROSTRIP ANTENNAS

Microstrip antennas are one of the most versatile types of antennas, suitable both as single elements in small devices and in large phased arrays. As a result, any current system that uses microstrip antennas has the potential to benefit from the addition of antenna reconfigurability.

Motivation for Reconfigurable Microstrip Antennas

In addition to their use in the large phased arrays described earlier, more and more microstrip antennas are found in portable/mobile devices for consumers. In these applications, reconfigurability can be implemented to select operating systems and reduce interference as well as to steer radiation patterns away from noise sources or users. Moreover, their well-known operational principles, illuminated through the cavity²⁵ and transmission line models,²⁶ lend themselves well to enable a wide range of reconfigurability. The examples here are confined to single element versions, but many of them can also be implemented in array formations as well.

Examples of Reconfigurable Microstrip Antennas

In work by Bhartia and Bahl,²⁷ the operating frequency range of a microstrip patch antenna was continuously tuned by using varactor diodes (varactors) at the radiating edges of the structure. The varactors operated with a reverse bias of between 0 and 30 volts that corresponded to capacitances of 2.4–0.4 pF. With a change in bias, the capacitance induced at the radiating edge is changed, resulting in a change in the effective electrical length of the patch. This allows for the continuous tuning of the operating frequency (though not typically of the instantaneous bandwidth) over a large band, which has been shown to be 20–30% depending on the type of microstrip antenna used.²⁷

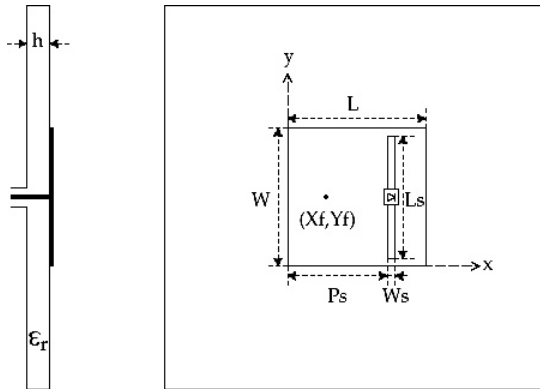


FIGURE 31-11 Geometry of a dual-frequency microstrip patch antenna with a switchable slot (PASS) (after F. Yang and Y. Ramat-Samii²⁸ © *Microwave and Optical Technology Letters* 2001)

Another way to affect discrete changes in a microstrip antenna's electrical length is to change the path of radiating currents without changing the overall footprint of the antenna. An example of this approach was proposed by Yang and Rahmat-Samii for a microstrip antenna.²⁸ Starting with a standard rectangular microstrip antenna, a slot is etched in the patch so that it is perpendicular to the direction of the main current of the patch's first resonance, as shown in Figure 31-11.²⁸ A PIN diode is positioned in the center of the slot to change the current paths on the patch. If the diode is off, then currents travel around the slot and the antenna operates in a lower frequency band, while when the diode is on, the effective length of the patch is shorter and the antenna operates in a higher frequency band. The frequency ratio between the upper and lower operating frequencies is controlled by the length of the slot, and as long as the slot length is not too long, the radiation pattern of the original antenna is largely preserved.²⁸ A similar approach can also be used to reconfigure the polarization of microstrip antennas.²⁹

While many reconfigurable designs are based on patches such as these, other shapes and modes for microstrip antennas have been used to deliver different kinds of radiation pattern reconfigurability. For instance, one design for a square spiral microstrip antenna uses switches placed at particular positions in order to deliver a broadside or 45 degree tilted beam over a common frequency bandwidth.³⁰ The basic antenna structure consists of a single turn square spiral antenna on an electrically thin substrate, shown in Figure 31-12,³⁰ which exhibits a broadside radiation pattern in its fundamental state. To enable reconfiguration, the antenna is equipped with two switched connections: the first between the spiral and ground, as indicated by N (not grounded) and G (grounded) and a second in a small gap in the spiral arm as indicated by S (short) and O (open).³⁰ When the first switch connects the line to ground and the second switch opens the in-line gap, a 45° tilt from broadside in radiation pattern results. Additionally, with activation of only the second switch, this new configuration provides a broadside radiation pattern at a higher operating frequency of 6.0 GHz.³⁰ Others have also studied similar spiral structures that provide a wealth of possibilities for switched tilts in radiation patterns^{31–32} as well as examined reconfigurable behavior when the antennas are integrated into packages.³³ Antennas based on a similar structure have also delivered broadside to endfire radiation characteristics with RF MEMS switched connections.³⁴

In another example, Zhang et al³⁵ developed a microstrip pattern reconfigurable antenna based on parasitic element tuning. The antenna, shown in Figure 31-13³⁵ is composed of

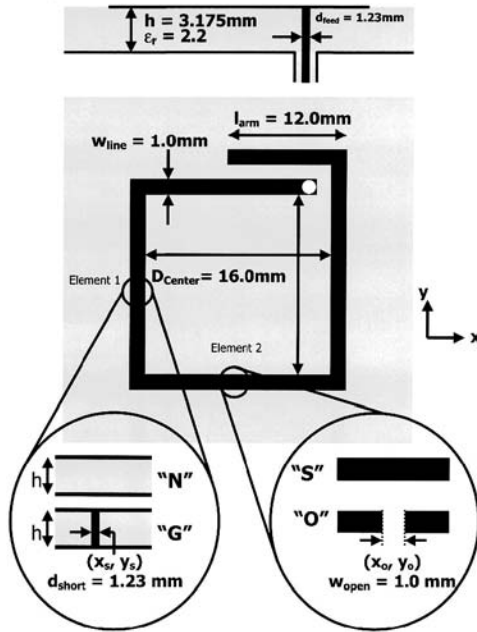


FIGURE 31-12 Radiation pattern and frequency reconfigurable square spiral microstrip antenna showing switch placements necessary for reconfiguration (after G. H. Huff et al³⁰ © IEEE 2003)

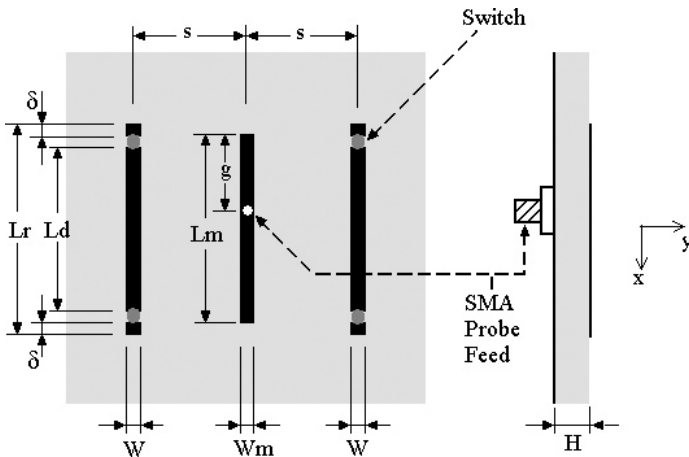


FIGURE 31-13 Physical structure and parameters of the reconfigurable microstrip parasitic array. Switching or tuning of parasitic elements on either side of the driven element provides beam tilt capabilities over a common impedance bandwidth (after S. Zhang³⁵ © IEEE 2004).

a single linear element with two parasitic elements positioned parallel to the driven element. Changes in the parasitic element lengths, achieved either with electronic switches (for example, diodes, FETs, and RF MEMS switches)³⁵ or varactors,³⁶ change the magnitudes and phases of the currents on the parasitic elements relative to the driven element. Tilts in the main beam in one plane can then be switched³⁵ or swept³⁶ as the lengths of the parasitic elements are changed. This antenna, as with all parasitically tuned antennas, can be analyzed theoretically using coupling and transmission line models.³⁷

Since microstrip designs typically include a substrate material other than air, possibilities exist, in theory, for reconfiguring antenna behavior through material changes. One example³⁸ of a frequency-tuned ferrite-based patch antenna is presented, which provides a 40% continuous tuning range with the variable static magnetic field in the plane of the substrate and perpendicular to the resonant dimension of the patch. However, the radiation performance of the design left much to be desired, with cross-polarization levels that were significantly higher than those expected from a traditional rectangular microstrip antenna.³⁸

Others have also investigated the properties of ferrite-based microstrip antennas,^{39–40} with results indicating that factors including non-uniform bias fields, as well as the multiple modes and resulting field distributions excited in a bulk ferrite substrate, may preclude their use in practical applications. One exception to this could be a design for a polarization-tunable microstrip antenna. This antenna is based on static magnetic biasing of a ferrite film, which takes advantage of the frequency tuning of the cross-polarized component of the radiated field to deliver tunable changes in polarization.⁴¹ Frequency-tunable patch designs based on the use of ferroelectric substrates or thin films, which typically have much higher relative permittivities than traditional low loss substrates, also face similar practical challenges.

31.6 RECONFIGURABLE SLOT ANTENNAS

Slot antennas, in their roles as the complements of dipole antennas, find applications in a number of systems that require low profile, wide bandwidths, and bi-directional radiation patterns. It is straightforward to integrate slot antennas into existing conductive structures or ground planes without additional structural support.

Motivation for Reconfigurable Slot Antennas

Typically, reconfigurable slot antennas are used to enable frequency agility in a system, since they are straightforward to design, simple to construct, and compatible with a number of switch topologies. Examples of reconfigurable slot antennas described next, some of which cover four or more operating bands, illustrate not only fundamental design principles but also strategies for size reduction and direct switch integration.

Examples of Reconfigurable Slot Antennas

Switched frequency radiating slots with a variety of geometries and radiating properties have been proposed by a number of researchers. Gupta et al⁴² designed a frequency-reconfigurable compact rectangular ring slot antenna that was fed with a single slot-line or coplanar waveguide line. Figure 31-14 shows a photograph of the antenna.⁴² Reconfigured with a total of eight PIN diode switches, the lower of the two frequencies is

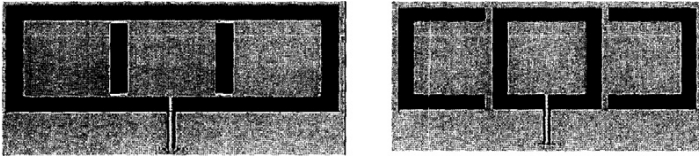


FIGURE 31-14 Photograph of a reconfigurable rectangular slot antenna in two different configurations that produce operation at two separate frequencies (after K.C. Gupta et al⁴² © IEEE 2000)

set by the perimeter of the outer loop. The upper frequency is set by the perimeter of the inner loop that was formed when shorter slot sections are activated in place of portions of the outer loop. Peroulis et al⁴³ developed a tunable antenna using four PIN diode switches that changed the effective length of the slot so it could operate in one of four different frequency bands between 530 MHz and 890 MHz without the need for changes in the matching network. A coplanar, waveguide-fed slot antenna utilizing similar principles for frequency tuning with PIN diodes is shown in Figure 31-15.⁴⁴ Often, the available switching technology can affect the choice of antenna geometry. For example,⁴⁵ a hybrid coplanar waveguide slot antenna with a 22% effective bandwidth (shown in Figure 31-16)⁴⁵ was designed to directly accommodate a particular RF MEMS switch geometry, making final fabrication and integration more straightforward.

Using tunable reactive loading, rather than switches, can result in continuously tuned slot antennas. In one case,⁴⁶ a one-wavelength slot antenna loaded with two one-port reactive FET components was tuned continuously. By changing the bias voltage, the reactances of the FETs were varied and the effective length of the slot was changed, thereby changing its operating frequency. The range of tuning varied over roughly 1 GHz centered on 10 GHz, producing a 10% tuning band. For this relatively small tuning range, the radiation patterns were essentially unchanged.⁴⁶ Similar tunable slot antennas loaded with varactors^{47–48} take advantage of higher order resonances to create tunable dual-band performance. Using a transmission line model of the loaded slot resonator, the position of the tuning elements can be determined to enable relatively independent tuning of the two bands.⁴⁸

Slot antennas have also been implemented for the purpose of delivering polarization reconfigurability. In one case, the slot-ring antenna⁴⁹ uses PIN diodes to reconfigure between linear and circular polarization or between two circular polarizations. The general topology

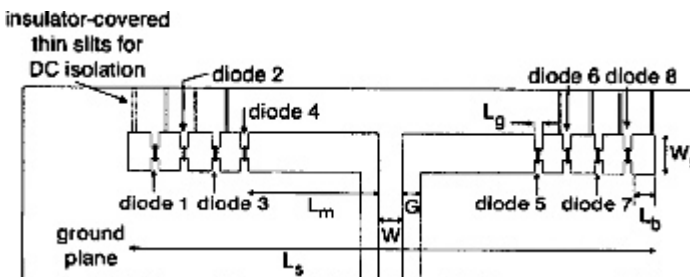


FIGURE 31-15 Frequency reconfigurable coplanar waveguide-fed slot antenna with eight diode switches and associated bias lines (after J. M. Laheurte⁴⁴ © IET 2001)

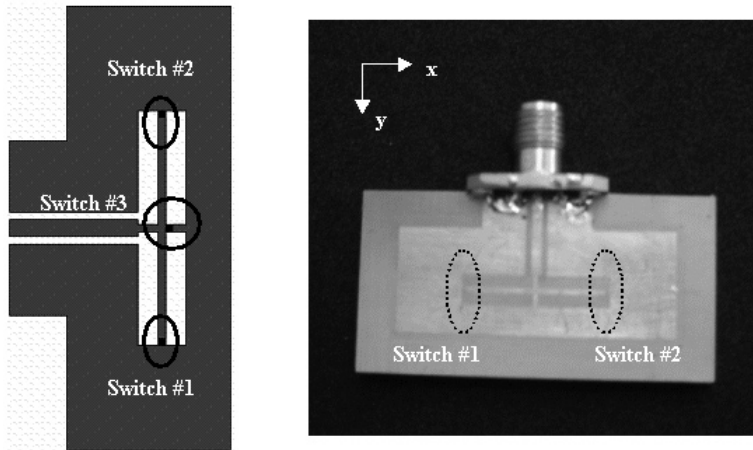


FIGURE 31-16 Reconfigurable hybrid folded slot/slot dipole antenna: antenna geometry including all three switch locations (left) and photograph of the fabricated antenna using only two of the three switches (right) (after G. H. Huff and J. T. Bernhard⁴⁵ © IEEE 2005)

of the antenna is shown in Figure 31-17,⁴⁹ with the specific diode positions, biasing, and ground plane configurations for both designs shown in Figure 31-18.⁴⁹ For the linear-circular design shown in Figure 31-18a, forward-biasing the diodes across the small discontinuities at 45° and -135° (relative to the feed point) produces linear polarization, while reverse biasing the diodes delivers circular polarization.⁴⁹ The design in Figure 31-18b goes one step further and adds additional symmetric discontinuities to support switching between left- and right-handed circular polarizations.⁴⁹ In both designs, care is taken to divide the ground planes carefully to support proper DC biasing for the diodes while providing RF continuity through capacitors connected between ground plane sections. This antenna is a good example of the additional factors that must be taken into account when transitioning from a fixed to a reconfigurable antenna—the fundamental structure may remain the same, but critical adjustments are required to enable proper DC connections and RF performance.

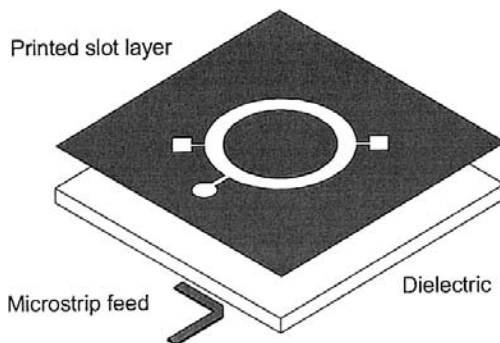


FIGURE 31-17 Basic topology of a microstrip-fed circularly polarized slot ring antenna (after M. K. Fries et al⁴⁹ © IEEE 2003)

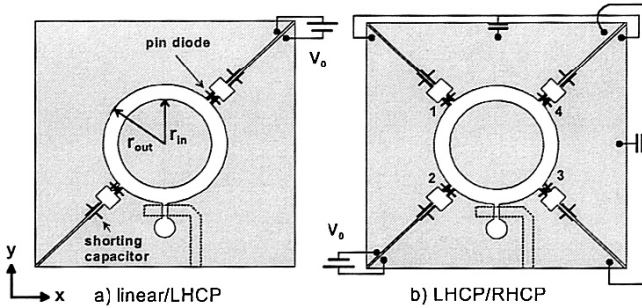


FIGURE 31-18 Two polarization reconfigurable slot ring antennas: (a) switchable between linear and left hand circular polarization and (b) switchable between left and right hand circular polarization (after M. K. Fries et al⁴⁹ © IEEE 2003)

31.7 RECONFIGURABLE MONOPOLE/DIPOLE ANTENNAS

Used in everything from phased arrays to portable communication devices, monopole and dipole antennas are both elemental and ubiquitous. While the theory explaining their operation is well known and based on a simple resonance principle, their application in specific situations typically requires adjustments to their length and packaging in order to deliver operation in one particular band.

Motivation for Reconfigurable Monopole/Dipole Antennas

When multifunctional systems require operation at multiple frequencies, adjustments to the basic monopole/dipole design can be accomplished either by creating a multiband structure, again based on fundamental or higher order resonances, or by making the antenna frequency reconfigurable. In some cases, the multiband approach is the more sensible within cost and complexity constraints, but frequency reconfigurable monopole/dipole antennas have roles to play where size, cost, noise immunity, security, and low observability are important factors. The reconfigurable monopoles and dipoles discussed here change the effective resonant length of the antenna by adding or removing part of the antenna length through electronic, optical, mechanical, or other means.

Examples of Reconfigurable Monopole/Dipole Antennas

Different kinds of switches (for example, PIN diodes, FETs, optical switches, and RF MEMS) have been implemented by several groups in frequency-tunable dipole and monopole antennas for various frequency bands. Roscoe et al implemented four PIN diodes in a printed dipole antenna to deliver three operating bands between 5.2 and 5.8 GHz.⁵⁰ For this specific design, the researchers limited the reconfiguration range so that a single feed network could be used for all of the operating bands. Kiriazi et al. present an example of a frequency reconfigurable antenna using RF MEMS switches.⁵¹ In this work, opening and closing a pair of RF MEMS switches reconfigures a simple dipole antenna printed on a high resistivity silicon substrate, allowing it to operate in one of two frequency bands. Others have applied the same approach to develop frequency reconfigurable Yagi antennas.⁵²

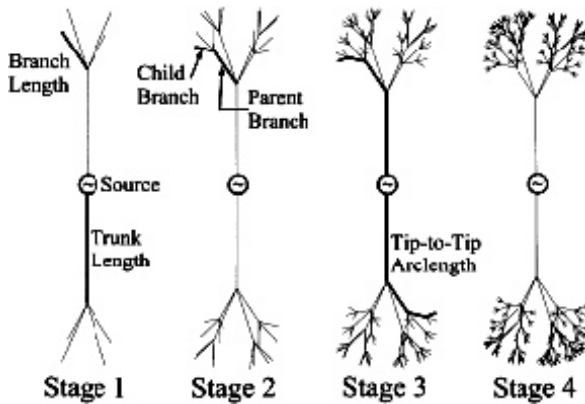


FIGURE 31-19 Reconfigurable three dimensional fractal tree dipole antennas with four branch levels (after J. S. Petko and D. H. Werner⁵³ © IEEE 2004)

More complicated structures based on fractal shapes, but sharing the same underlying principle, have been studied by numerous researchers. For instance, a three-dimensional fractal tree structure is proposed for use with either passive frequency traps or RF MEMS switches to deliver operation over multiple frequency bands.⁵³ An example of the evolution of the structure is shown in Figure 31-19.⁵³ For this antenna, the design methodology can produce operation over multiple contiguous bands, which is only limited by the practicality of implementing the necessary number of switches. Anagnostou et al. propose a reconfigurable monopole based on a Sierpinski gasket, with RF MEMS switches for connecting different sections of the antenna together in order to provide multiple operating bands.⁵⁴ Subsequent work, with the direct integration of RF MEMS switches in a dipole design and shown in Figure 31-20,⁵⁵ provided three separate operating bands with similar omnidirectional radiation characteristics.⁵⁵ Direct fabrication of the RF MEMS switches during the fabrication of the antenna helps to minimize package parasitics that may detune or limit the frequency behavior of the antenna.

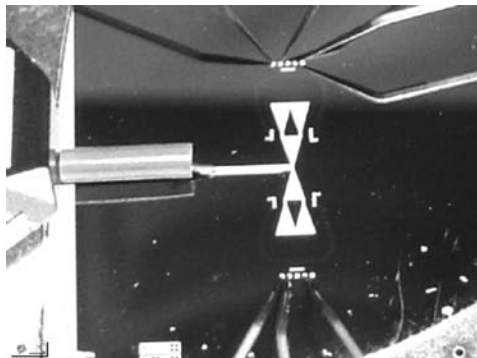


FIGURE 31-20 Photograph of a reconfigurable planar Sierpinski dipole antenna with integrated RF MEMS switches (after D. E. Anagnostou et al⁵⁵ © IEEE 2006)

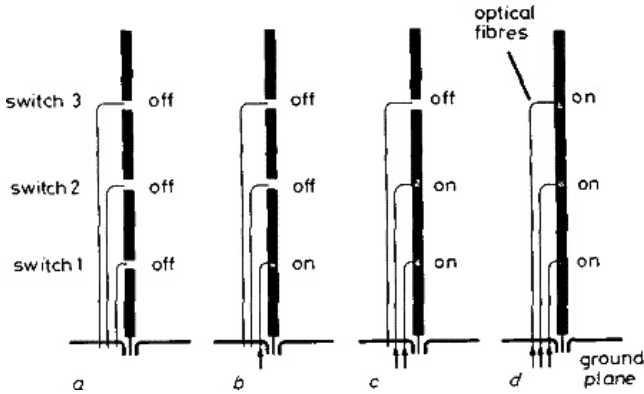


FIGURE 31-21 Diagram of a reconfigurable monopole antenna switchable between four frequency bands with optoelectronic switches (after J. L. Freeman et al⁵⁶ © IET 1992)

Using the same concept but different switch technologies, Freeman et al changed the effective length of a monopole antenna using optical switches. This helped to eliminate many of the switch and bias line effects that can occur with other kinds of switches.⁵⁶ A diagram of this antenna is shown in Figure 31-21.⁵⁶

An excellent example of reconfigurable dipole antennas improving system performance is presented by Piazza and Dandekar.⁵⁷ In this work, reconfigurable printed dipole arrays using PIN diodes have been implemented to enable multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems.⁵⁷ For this application, a set of two printed frequency reconfigurable dipoles were spaced a quarter wavelength apart at 2.45 GHz. They were then included at both the transmitter and receiver of a MIMO communication system, shown in Figure 31-22.⁵⁷ In the short configuration of the antennas,

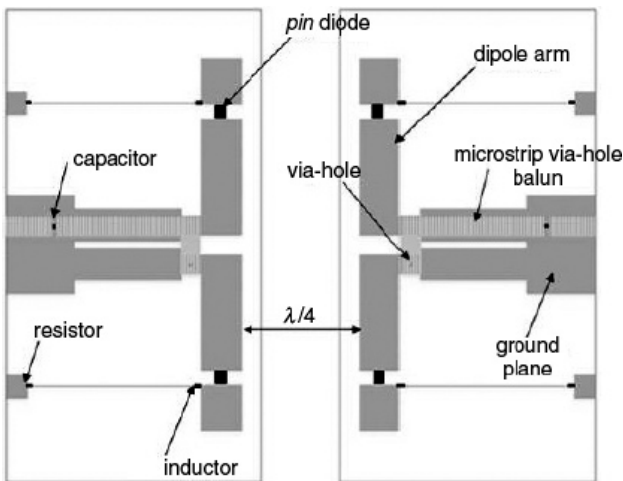


FIGURE 31-22 Diagram of an array of two reconfigurable dipoles used to achieve pattern diversity at a single operating frequency in MIMO systems (after D. Piazza and K. R. Dandekar⁵⁷ © IET 2006)

the elements operated as half-wavelength dipoles. In the long configuration, the antennas were driven in a higher order mode where they were three quarter wavelengths long at the same frequency of 2.45 GHz. By changing the lengths of the antennas but by keeping the operating frequency the same, the researchers were able to achieve four different radiation patterns with each of the arrays, thereby providing a degree of pattern reconfigurability that helped to improve MIMO system capacity.⁵⁷

It is worthwhile to note that implementation of switches, in both monopoles and dipoles, often prevents the antenna from being a self-supporting structure. The need for surfaces for switch and bias line connections usually requires some kind of planar or other dielectric support, which may distort radiation patterns and increase antenna loss and weight slightly.

31.8 CONCLUDING REMARKS

Antenna reconfigurability itself isn't new, but the opportunities for integrating reconfigurable antennas into systems that can fully exploit antenna capabilities are. These opportunities include software defined radio, cognitive radio, MIMO systems, multifunction consumer wireless devices, and high performance phased arrays. As all of the examples discussed in this chapter demonstrate, new antenna designs can be developed to change their frequencies, bandwidths, polarizations, and radiation patterns. They can be designed to solve specific problems now as well as to make systems more flexible and responsive to future operational requirements. However, the challenges for the future of this area of antenna engineering do not lie only in antenna design. Two other significant areas for future work include the development of appropriate control networks and the wide adoption of antenna reconfigurability as an important feature in overall system design. While initial efforts to demonstrate the benefits of reconfigurability in a system-level context have been made by researchers such as Piazza and Dandekar,⁵⁷ a great deal remains to be done. As this work progresses, these new capabilities can be used effectively to deliver high throughput, high reliability, and high security wireless connections in the future.

REFERENCES

1. P. J. B. Clarricoats and H. Zhou, "The Design and Performance of a Reconfigurable Mesh Reflector Antenna," *Proc. IEE Seventh Int. Conf. on Antennas and Propagat.*, vol. 1 (April 1991): 322–325.
2. P. J. B. Clarricoats, H. Zhou, and A. Monk, "Electronically Controlled Reconfigurable Reflector Antenna," *Proc. IEEE Int. Symp on Antennas and Propagation*, vol. 1 (June 1991): 179–181.
3. G. Washington, H. S. Yoon, M. Angelino, and W. H. Theunissen, "Design, Modeling, and Optimization of Mechanically Reconfigurable Aperture Antennas," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 5 (May 2002): 628–637.
4. L. N. Pringle, P. H. Harms, S. P. Blalock, G. N. Kiesel, E. J. Kuster, P. G. Friederich, R. J. Prado, J. M. Morris, and G. S. Smith, "A Reconfigurable Aperture Antenna Based on Switched Links Between Electrically Small Metallic Patches," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 6 (June 2004): 1434–1445.
5. A. Fathy, A. Rosen, H. Owen, F. McGinty, D. McGee, G. Taylor, R. Amantea, P. Swain, S. Perlow, and M. ElSherbiny, "Silicon-Based Reconfigurable Antennas—Concepts, Analysis, Implementation, and Feasibility," *IEEE Trans. Microwave Theory and Techn.*, vol. 51, no. 6 (June 2003): 1650–1661.
6. M. ElSherbiny, A. E. Fathy, A. Rosen, G. Ayers, and S. M. Perlow, "Holographic Antenna Concept, Analysis and Parameters," *IEEE Trans. Antennas Propag.*, vol. 52, no. 3 (March 2004): 830–839.

7. K. Chang, M. Li, T-Y. Yun, and C. T. Rodenbeck, "Novel Low-Cost Beam Steering Techniques," *IEEE Trans. Antennas Propag.*, vol. 50, no. 5 (May 2002): 618–627.
8. C. T. Rodenbeck, M. Li, and K. Chang, "Design and Analysis of a Reconfigurable Dual-Beam Grating Antenna for Low-Cost Millimeter-Wave Beam-Steering," *IEEE Trans. Antennas Propag.*, vol. 52, no. 4 (April 2004): 999–1006.
9. W. Rotman and A. Maestri, "An Electromechanically Scannable Trough Waveguide Array," *Proc. IRE International Convention Record*, vol. 8 (March 1960): 67–83.
10. W. Rotman and A. Oliner, "Asymmetrical Trough Waveguide Antennas," *IEEE Trans. Antennas Propag.*, vol. 7 (April 1959): 153–162.
11. G. H. Huff and J. T. Bernhard, "Electromechanical Beam Steering of a Trough Waveguide Antenna Using Cantilever Perturbations," *Proc. 2005 Antenna Applications Symposium* (September 2005): 152–165.
12. D. Stevenpiper, J. Schaffner, J. J. Lee, and S. Livingston, "A Steerable Leaky Wave Antenna Using a Tunable Impedance Ground Plane," *IEEE Antennas and Wireless Propagation Letters*, vol. 1 (2002): 179–182.
13. Y. E. Erdemli, K. Sertel, R. A. Gilbert, D. E. Wright, and J. L. Volakis, "Frequency-Selective Surfaces to Enhance Performance of Broad-Band Reconfigurable Arrays," *IEEE Trans. Antennas Propag.*, vol. 50 (December 2002): 1716–1724.
14. Y. E. Erdemli, R. A. Gilbert, and J. L. Volakis, "A Reconfigurable Slot Aperture Design Over a Broad-Band Substrate/Feed Structure," *IEEE Trans. Antennas Propag.*, vol. 52 (November 2004): 2860–2870.
15. G. H. Huff, J. Feng, S. Zhang, and J. T. Bernhard, "Behavior of Pattern and/or Frequency Reconfigurable Antennas in Small Arrays," *Proc. 2003 IEEE/URSI Int. Symp. on Antennas and Propagation*, URSI (June 2003): 151.
16. G. H. Huff, J. Feng, and J. T. Bernhard, "A Small Array of Broadside to Endfire Radiation Reconfigurable Antennas," *Proc. 2003 Antenna Applications Symposium* (Sept. 2003): 147–161.
17. K. Hietpas, G. H. Huff, and J. T. Bernhard, "Investigation of Phased Array Beamsteering Using Reconfigurable Antennas," *Proc. IASTED Int. Conf. on Antennas, Radar, and Wave Propagation* (July 2004): 41–44.
18. G. H. Huff, "A Radiation Reconfigurable Magnetic Line Source Antenna: Modeling, Integration with RF MEMS, and Applications," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2006.
19. R. F. Harrington, "Reactively Controlled Directive Arrays," *IEEE Trans. Antennas Propag.*, vol. 26 (May 1978): 390–395.
20. J. Luzwick and R. Harrington, "A Reactively Loaded Aperture Antenna Array," *IEEE Trans. Antennas Propag.*, vol. 26, no. 4 (July 1978): 543–547.
21. D. V. Thiel, "Switched Parasitic Antennas and Controlled Reactance Parasitic Antennas: A Systems Comparison," *Proc. IEEE Antennas and Propagation Int. Symp.*, vol. 3 (June 2004): 3211–3214.
22. R. Schlub, D. V. Thiel, J. W. Lu, and S. G. O'Keefe, "Dual-Band Six-Element Switched Parasitic Array for Smart Antenna Cellular Communications," *Electron. Lett.*, vol. 36, no. 16 (August 2000): 1342–1343.
23. R. Schlub, J. Lu, and T. Ohira, "Seven-Element Ground Skirt Monopole ESPAR Antenna Design from a Genetic Algorithm and Finite Element Method," *IEEE Trans. Antennas Propag.*, vol. 51, no. 11 (November 2003): 3033–3039.
24. T. Ohira and K. Gyoda, "Hand-Held Microwave Direction-of-Arrival Finder Based on Varactor-Tuned Analog Aerial Beamforming," *Proc. IEEE Asia Pacific Conference* (Taipei), vol. 2 (December 2001): 585–588.
25. Y. T. Lo, D. Solomon, and W. Richards, "Theory and Experiment on Microstrip Antennas," *IEEE Trans. Antennas Propag.*, vol. 27, no. 2 (March 1979): 137–145.
26. A. Derneryd, "Linearly Polarized Microstrip Antennas," *IEEE Trans. Antennas Propag.*, vol. 24, no. 6 (November 1976): 846–851.

27. P. Bhartia and I. J. Bahl, "Frequency Agile Microstrip Antennas," *Microwave J.*, vol. 25 (October 1982): 67–70.
28. F. Yang and Y. Rahmat-Samii, "Patch Antenna with Switchable Slot (PASS): Dual Frequency Operation," *Microwave and Optical Technology Letters*, vol. 31 (November 2001): 165–168.
29. F. Yang and Y. Rahmat-Samii, "A Reconfigurable Patch Antenna Using Switchable Slots for Circular Polarization Diversity," *IEEE Microwave and Wireless Components Lett.*, vol. 12 (March 2002): 96–98.
30. G. H. Huff, J. Feng, S. Zhang, and J. T. Bernhard, "A Novel Radiation Pattern and Frequency Reconfigurable Single Turn Square Spiral Microstrip Antenna," *IEEE Microwave and Wireless Components Lett.*, vol. 13, no. 2 (February 2003): 57–59.
31. C. Jung, M. Lee, G. P. Li, and F. De Flaviis, "Reconfigurable Scan-Beam Single-Arm Spiral Antenna Integrated with RF-MEMS Switches," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2 (February 2006): 455–463.
32. A. Mehta, D. Mirshekar-Syahkal, and H. Nakano, "Beam Adaptive Single Arm Rectangular Spiral Antenna with Switches," *IEE Proceedings-Microwaves, Antennas and Propagation*, vol. 153, no. 1 (February 2006): 13–18.
33. G. H. Huff, J. Feng, S. Zhang, and J. T. Bernhard, "Directional Reconfigurable Antennas on Laptop Computers: Simulation, Measurement, and Evaluation of Candidate Integration Positions," *IEEE Trans. Antennas Propag.*, vol. 52 (December 2004): 3220–3227.
34. G. H. Huff and J. T. Bernhard, "Integration of Packaged RF MEMS Switches with Radiation Pattern Reconfigurable Square Spiral Microstrip Antennas," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2 (February 2006): 464–469.
35. S. Zhang, G. H. Huff, J. Feng, and J. T. Bernhard, "A Pattern Reconfigurable Microstrip Parasitic Array," *IEEE Trans. Antennas Propag.*, vol. 52 (October 2004): 2773–2776.
36. S. Zhang, G. Huff, G. Cung, and J. T. Bernhard, "Three Variations of a Pattern Reconfigurable Microstrip Parasitic Array," *Microwave and Optical Technology Letters*, vol. 45 (June 2005): 369–372.
37. S. Zhang, "A Pattern Reconfigurable Microstrip Parasitic Array: Theory, Design, and Applications," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2005.
38. D. M. Pozar and V. Sanchez, "Magnetic Tuning of a Microstrip Antenna on a Ferrite Substrate," *Electron. Lett.*, vol. 24 (June 1988): 729–731.
39. R. K. Mishra, S. S. Pattnaik, and N. Das, "Tuning of Microstrip Antenna on Ferrite Substrate," *IEEE Trans. Antennas Propag.*, vol. 41, no. 2 (February 1993): 230–233.
40. A. D. Brown, J. L. Volakis, L. C. Kempel, and Y. Botros, "Patch Antennas on Ferromagnetic Substrates," *IEEE Trans. Antennas Propag.*, vol. 47 (January 1999): 26–32.
41. P. Rainville and F. Harackiewicz, "Magnetic Tuning of a Microstrip Patch Antenna Fabricated on a Ferrite Film," *IEEE Microwave and Guided Wave Lett.*, vol. 2 (December 1992): 483–485.
42. K. C. Gupta, J. Li, R. Ramadoss, and C. Wang, "Design of Frequency-Reconfigurable Slot Ring Antennas," *Proc. IEEE/URSI Int. Symp. on Antennas and Propagation*, vol. 1 (July 2000): 326.
43. D. Peroulis, K. Sarabandi, and L. P. B. Katehi, "Design of Reconfigurable Slot Antennas," *IEEE Trans. Antennas Propag.*, vol. 53 (February 2005): 645–654.
44. J. M. Laheurte, "Switchable CPW-fed Slot Antenna for Multifrequency Operation," *Electronics Letters*, vol. 37, no. 25 (December 2001): 1498–1500.
45. G. H. Huff and J. T. Bernhard, "Frequency Reconfigurable CPW-fed Hybrid Folded Slot/Slot Dipole Antenna," *Proc. IEEE/ACES Int. Conf. on Wireless Communications and Applied Computational Electromagnetics* (April 2005): 574–577.
46. S. Kawasaki and T. Itoh, "A Slot Antenna with Electronically Tunable Length," *Proc. IEEE/URSI Int. Symp. on Antennas and Propagation*, vol. 1 (June 1991): 130–133.
47. N. Behdad and K. Sarabandi, "A Varactor-Tuned Dual-Band Slot Antenna," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2 (February 2006): 401–408.
48. N. Behdad and K. Sarabandi, "Dual-Band Reconfigurable Antenna with a Very Wide Tunability Range," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2 (February 2006): 409–416.

49. M. K. Fries, M. Grani, and R. Vahldieck, "A Reconfigurable Slot Antenna with Switchable Polarization," *IEEE Microwave and Wireless Components Letters*, vol. 13, no. 11 (November 2003): 490–492.
50. D. J. Roscoe, L. Shafai, A. Ittipiboon, M. Cuhaci, and R. Douville, "Tunable Dipole Antennas," *Proc. IEEE Int. Symp. on Antennas and Propagation*, vol. 2 (July 1993): 672–675.
51. J. Kiriazi, H. Ghali, H. Radaie, and H. Haddara, "Reconfigurable Dual-Band Dipole Antenna on Silicon using Series MEMS Switches," *Proc. IEEE Int. Symp. on Antennas and Propagation*, vol. 1 (June 2003): 403–406.
52. M. A. Ali and P. Wahid, "A Reconfigurable Yagi Array for Wireless Applications," *Proc. IEEE/URSI Int. Symp on Antennas and Propagation*, vol. 1 (2002): 466–468.
53. J. S. Petko and D. H. Werner, "Miniature Reconfigurable Three-Dimensional Fractal Tree Antennas," *IEEE Trans. Antennas Propag.*, vol. 52, no. 8 (August 2004): 1945–1956.
54. D. Anagnostou, M. T. Chryssomallis, J. C. Lyke, and C. G. Christodoulou, "Re-configurable Sierpinski Gasket Antenna using RF-MEMS Switches," *Proc. IEEE Int. Symp. on Antennas and Propagation*, vol. 1 (June 2003): 375–378.
55. D. E. Anagnostou, G. Zheng, M. T. Chryssomallis, J. C. Lyke, G. E. Ponchak, J. Papapolymerou, and C. G. Christodoulou, "Design, Fabrication, and Measurements of an RF-MEMS-based Self-Similar Reconfigurable Antenna," *IEEE Trans. Antennas Propag.*, vol. 54, no. 2 (February 2006) 422–432.
56. J. L. Freeman, B. J. Lamberty, and G. S. Andrews, "Optoelectronically Reconfigurable Monopole Antenna," *Electron. Lett.*, vol. 28, no. 16 (July 1992): 1502–1503.
57. D. Piazza and K. R. Dandekar, "Reconfigurable Antenna Solution for MIMO-OFDM Systems," *Electron. Lett.*, vol. 42, no. 8 (April 2006): 446–447.

Chapter 32

Active Antennas

Zoya Popović
Nestor Lopez
Sebastien Rondineau

University of Colorado at Boulder

CONTENTS

32.1	INTRODUCTION AND BASIC TERMS	32-2
32.2	FREQUENCY-AGILE ACTIVE ANTENNAS	32-9
32.3	OSCILLATOR ANTENNAS AND ARRAYS	32-11
32.4	AMPLIFIER ANTENNAS AND ARRAYS	32-17
32.5	FREQUENCY-CONVERTING ANTENNAS	32-23
32.6	RECTENNAS	32-26
32.7	OTHER ACTIVE ANTENNAS	32-30

32.1 INTRODUCTION AND BASIC TERMS

Antennas are used either in transmission, in which case they are fed from the output of some amplifier, or in reception, in which case they provide the RF input to an LNA. Therefore, in some sense, all antennas are active. In the context of this chapter, however, an *active antenna* implies an antenna integrated intimately with an active circuit, including the DC bias circuit, and without an isolator or circulator between them. The absence of isolator/circulator implies that neither the antenna nor the circuit needs to be designed in a 50- Ω environment. Figure 32-1 shows block diagrams of several types of active antennas, classified according to their functionality. Note that this chapter does not discuss in detail phased array antennas. This is a separate and large topic in itself, with excellent overviews given in the literature^{1,2} as well as in papers on T/R modules.³

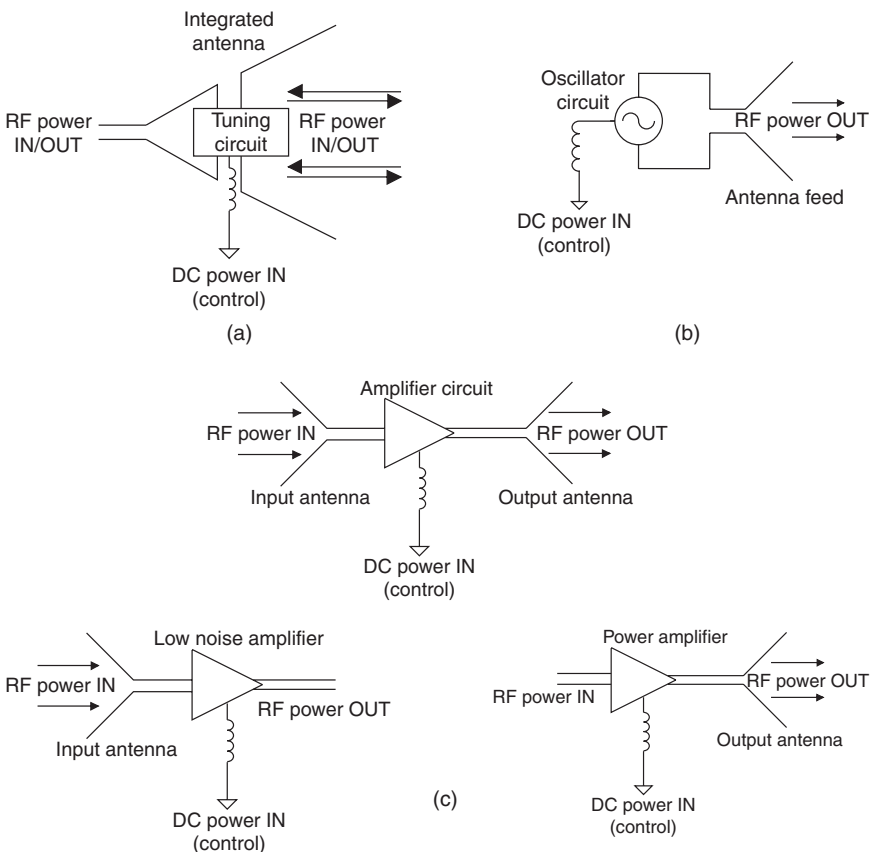


FIGURE 32-1 Schematic diagrams of a few examples of active antennas: (a) actively tuned antenna, (b) oscillator antenna, and (c) amplifier antenna with repeater, receiver, and transmitter subclasses. An *active antenna* implies an antenna integrated intimately with an active circuit.

- **Oscillator antennas** A two- or three-terminal negative-resistance device can be connected directly to the terminals of a single antenna element or an array of elements. In this case, DC power is converted to radiated RF power. An example of a patch antenna in the feedback loop of a transistor, similar to the one in Chang et al.,⁶ is shown in Figure 32-2b. Oscillator antennas have been discussed for applications such as low-cost sensors,⁷ power combining,^{8,9,10} and synchronized scanning antenna arrays.¹¹
- **Amplifier antennas** An active device is connected to the terminals of an antenna element to provide amplification in receive mode or transmit mode. In the former case, the matching between the antenna and active element usually optimizes noise, while in the latter case, the matching optimizes power and/or efficiency. Figure 32-2c shows an example of a repeater element with two slot antennas and a prematched amplifier chip, similar to the work in Hollung et al.¹² In this case, increase in gain is enabled by adding DC power to the antenna, and it becomes difficult to separate antenna gain from circuit gain. Amplifier antennas find applications in transmitters where spatial power combining can be achieved with an array, and in receivers where the feedline loss, which contributes to the total noise figure, can be eliminated by directly connecting an LNA to the receiving antenna.
- **Frequency-conversion antennas** A two- or three-terminal active device integrated with an antenna can provide direct down or up conversion of a radiated signal, at frequencies that are direct harmonics or subharmonics of a fundamental frequency (multipliers, dividers), or at frequencies with a given offset from the operating frequency (mixers). Figure 32-2d shows an example of a slot antenna with a Schottky diode, which can be used for frequency doubling since the slot is matched to the diode impedance at both the input frequency and its harmonic.¹³ Such antennas have applications in receivers,¹⁴ mixers with high dynamic range,¹⁵ detectors for millimeter-wave and THz receivers,¹⁶ phase-conjugating RFID type antennas,¹⁷ and high-frequency generation.¹⁸ A special case of frequency-conversion antennas is when a two- or three-terminal rectifying device is connected directly to the terminals of a receiving antenna in such a way that the received RF power is converted with optimal efficiency to DC power, while harmonic production and re-radiation is minimized. This type of active antenna is referred to as a *rectenna*. Such antennas have been applied to RFID tags,¹⁹ sensor powering for cases when there is no solar power and where it is difficult to replace batteries,²⁰ directed narrow-beam array power beaming,²¹ and for energy recycling and/or scavenging.²²
- **Other active antennas** A number of antennas in which the polarization properties (see e.g., Peroulis et al.²³) or radiation pattern can be varied based on control of an active device have been demonstrated. A separate chapter in this book covers reconfigurable antennas, and in this chapter we give a brief description of a few examples of such antennas, including an example of optically controlled antennas and arrays.²⁴

Figures of Merit for Active Antennas

Most standard antenna figures of merit can be applied to active antennas, but care must be taken when defining all properties that include power, since some of the power comes from external DC or RF sources. Properties that do not include power and are defined as in the case of passive antennas are

- Polarization (linear, circular, elliptical; described by an axial ratio)
- Bandwidth (usually defined through return loss, but it is important to also quantify radiation patterns and efficiency, i.e., gain vs. frequency)
- Normalized power radiation pattern

Properties that do include power, for which careful definitions are necessary, defined in the far field are

- Efficiency, defined as

$$\eta = \frac{P_{\text{radiated}}}{P_T + P_{\text{loss}}},$$

where the power loss in the antenna can actually be gain if only RF power is taken into account.

- Gain, defined as

$$G(\theta, \phi) = \eta \frac{S(r, \theta, \phi)}{S_i(r)} = \eta \frac{4\pi r^2 S(r, \theta, \phi)}{P_T}$$

The transmitted power P_T is not easy to determine when the antenna and active circuit are directly integrated. In such cases, the following quantity is relevant:

- Effective radiated power (ERP) with respect to an isotropic source (thus referred to sometimes as EIRP):

$$\text{ERP} = P_T G$$

measured for active transmitting antenna by measuring power density in the far field and assuming it is radiated by an isotropic radiator.

- Effective area, defined as

$$A(\theta, \phi) = \frac{P_R}{S(\theta, \phi)} = \frac{\lambda^2 \cdot G(\theta, \phi)}{4\pi}$$

where G could be the passive or active antenna gain. The passive gain is difficult to access in an integrated antenna, while the active gain yields an effective area that can be larger than the geometric area due to the gain of the active device, and at the expense of DC power.

- Power combining efficiency, relevant in the case of arrays of active transmitting and receiving antennas, and defined as

$$\text{PCE}_{\text{array}} = \frac{\text{ERP}}{G \cdot N \cdot P_i}$$

where the ERP is measured for the entire array, G is found from the geometric area of the array or measured pattern, N is the number of array elements, and P_i is the power of a single isolated element.

- Isotropic conversion loss of a frequency-conversion active antenna has been defined first in Stephan and Itoh²⁵ as

$$L_{\text{iso}} = \frac{P_{\text{IF}}}{P_{\text{RF}} G} = \frac{L}{G_{\text{ant}}}$$

where the measured quantities are the ERP and the down-converted power at the IF frequency, L is the conversion loss of the mixer, and G_{ant} the gain of the antenna. The meaning of this expression is that the same mixer, impedance matched to two different antennas, will have a higher IF output power when connected to an antenna with higher gain.

- Isotropic conversion gain of an oscillator antenna is defined as

$$G_{\text{iso}} = \frac{\text{ERP}}{P_{\text{DC}}} = \frac{P_{\text{RF}} G}{P_{\text{DC}}} = \eta G$$

where the measured quantities are the ERP and the input DC power. The antenna gain can be obtained by simulation or estimated from the measured radiation pattern, which then provides an estimate of the radiated RF power and efficiency of the oscillator circuit.

Modeling of Active Antennas and Limitations of Current CAD Tools

The most efficient way to model an active antenna is to start at the highest level of abstraction, which is a circuit simulator. The antenna can be included as a frequency-variable impedance obtained from a full-wave field simulation. This, however, does not give information about the antenna gain, polarization, or radiation pattern.

The design of an active antenna involves direct matching between the active circuit and antenna. Antennas have a complex impedance with both real and imaginary parts, which vary in frequency. The impedance of active devices is also frequency-dependent. Therefore, matching both the antenna and active device to a common $50\text{-}\Omega$ impedance is not necessarily optimal, although it is common practice due to the fact that the antenna is usually designed separately from the active circuit. An example of an amplifier antenna in which the impedance of the antenna is not $50\text{ }\Omega$, but is designed to have specific complex impedances at both the fundamental and second harmonic frequencies, is given in Weiss and Popovic.²⁶ The impedance of a nonresonant microstrip-circuit-fed antenna is in this case chosen to optimize transistor efficiency by wave-shaping of the voltage and currents.

Consider the Smith chart plots in Figure 32-3. The feedpoint reflection coefficient with respect to $50\text{ }\Omega$ for two antennas is plotted over a broad frequency range. One antenna is a linearly polarized patch centered at 10 GHz and plotted from 8 to 12 GHz in Figure 32-3a. The other is a self-complementary Archimedean spiral antenna that operates above 1.7 GHz and is simulated up to 18 GHz . In an integrated active antenna, the impedance of an antenna needs to be matched to that of the active device. Figure 32-3b shows the input and output reflection coefficients into $50\text{ }\Omega$ for a bipolar and FET transistor from 0.5 to 18 GHz , showing significant variation over frequency. Matching either antenna from Figure 32-3a to a transistor over a range of frequencies can be done in a number of ways.

As an illustration, consider a unit cell, shown in Figure 32-4a, of an amplifier array.²⁷ An off-center-fed $100\text{-}\Omega$ second-resonant slot antenna is the output of a two-stage LNA, while a folded slot receives the input wave in a configuration corresponding to that in Figure 32-1c (top). The off-center-fed second-resonance slot and the folded slot were designed using a method of moments code, where the off-center-fed slot result is shown in Figure 32-4b.

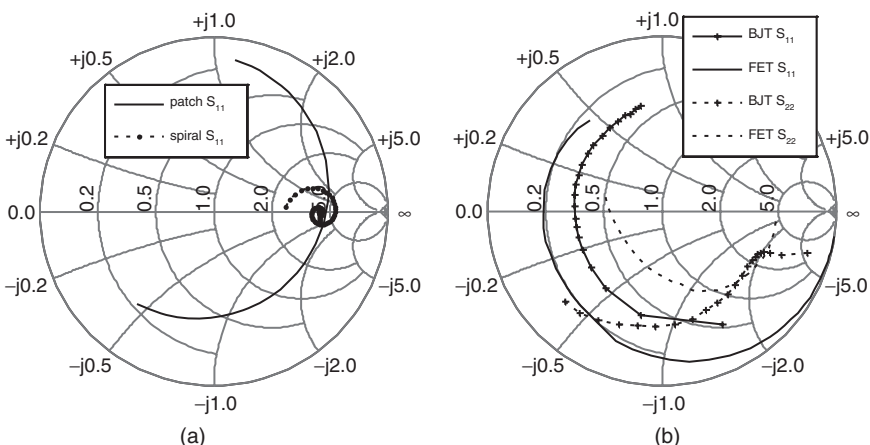


FIGURE 32-3 Smith chart plots of (a) antenna feedpoint reflection coefficient (with respect to $50\text{ }\Omega$) for a narrowband linearly polarized (CP) patch antenna plotted from 8 to 12 GHz , and a broadband CP spiral antenna (0.5 to 18 GHz); and (b) input and output S-parameter variation of a Silicon bipolar transistor (NE6800f) and a GaAs FET (NE7130Na) from 0.5 to 18 GHz . The BJT has gain up to around 10 GHz , while the FET operates to 18 GHz .

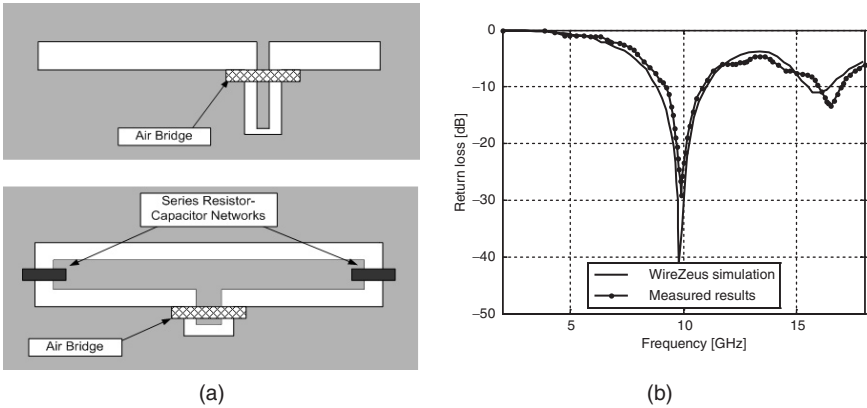


FIGURE 32-4 (a) An off-center-fed 100- Ω second-resonant slot antenna is the output of a two-stage LNA, while a folded slot receives the input wave in a configuration corresponding to that in Figure 32-1c. The printed slot antennas are designed using full-wave analysis. (b) Off-center-fed slot return loss.

The obtained frequency-variable impedances are then inserted into the Agilent ADS circuit simulator to check for stability. It was found that the amplifier was unstable, and a resistively loaded stabilization network that also serves as a bias network was added. Full-wave simulations were used not just for the antenna element design, but also to verify coupling inside a passive unit cell (with a thru-line instead of the transistor). Since the gain of the LNA is around 25 dB in this case, the coupling between the input and output orthogonally polarized antennas needs to be well below this value to ensure stability.

Alternatively, some electromagnetic field simulators have the option of including an S-parameter multiport network in the full-wave simulation. Figure 32-5 shows an example simulation of a transistor with a patch antenna at its output port, with a general connection block between the drain of the FET and the patch antenna. First, the FET from Figure 32-3b is directly connected to the square patch antenna at the radiating edge. Ideally, the

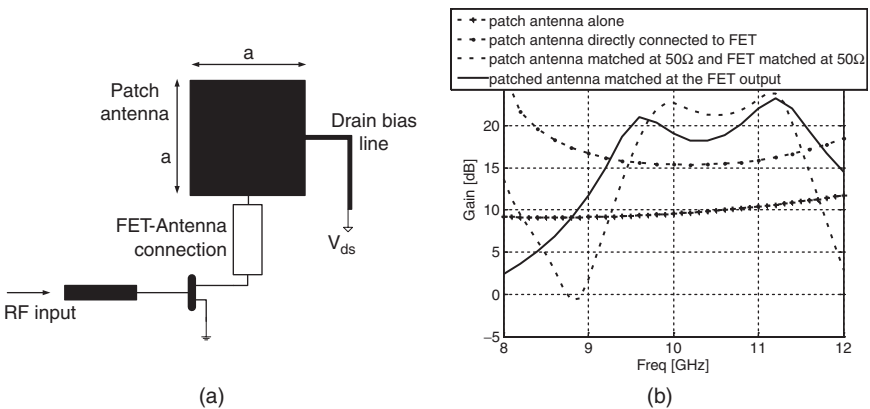


FIGURE 32-5 (a) Patch antenna matched to a transistor input impedance, with added bias line through the patch null. (b) Simulated active antenna gain for transistor directly connected to the patch at the radiating edge, both antenna and transistor matched to 50 Ω , and transistor directly matched to the antenna with a single nonoptimized matching section.

radiating-edge-fed square patch is linearly polarized with symmetric E- and H-plane patterns. The resulting active antenna gain, defined as $G_{AA} = \text{ERP} / P_{\text{RF-in}}$, is shown in Figure 32-5b. Then, both antenna and transistor are matched to 50Ω and the gain of the cascade is shown in the same plot. This is the standard design method and gives improved gain at the expense of space used by two matching networks. Finally, the transistor is directly matched to the antenna with a single matching section, and the gain is shown to change over frequency. For this particular nonoptimal match, the gain is lower than for the $50\text{-}\Omega$ case, and is used only as an illustration of the possible different outcomes. For these simulations, a method-of-moments analysis (Ansoft Nexxim) is performed with a two-port S -parameter (s2p) transistor “black box” for a specific bias point. A bias line is added to the field simulation through the RF null of the patch antenna, as shown in Figure 32-5a. There is no observable change in the active antenna gain, but the radiation pattern loses its symmetry in the H plane due to the addition of the bias line. The amount of pattern distortion depends on the exact position and thickness of the bias line. The E-plane pattern symmetry is affected by the transistor output port, and the polarization is slightly degraded.

This type of analysis gives far-field active antenna parameters, such as gain (defined in Figure 32-5 consistent with the IEEE definition), polarization, radiation pattern, as well as near-field quantities such as current distribution. Because the gain definition only takes into account RF power quantities, the simulated gain for an active antenna is larger than the passive antenna gain as the input DC power is not included. This type of analysis is more computationally intensive than circuit-based analysis.

A reasonable design procedure starts from an EM simulation of the antenna which gives a frequency-dependent impedance for inclusion into a circuit simulation. After the circuit is optimized using linear or nonlinear analysis, an equivalent multiport can be included in a final field simulation in order to validate potential effects of the circuit on radiation pattern.

Active Antenna Measurements

When characterizing the performance of active antennas, it is important to take into account the active circuit properties. An example active antenna measurement for the case of a receiving amplifier antenna of the type given in Figure 32-1c (bottom left) is given here. The only quantities that are accessible are the received RF power at the output of the LNA and the DC power input to the LNA. The received RF power is a function of frequency, magnitude, incidence angle, and polarization of the input wave. For an incidence angle (θ, ϕ) and a power P_T transmitted from an antenna in the far field at a distance r away, the power received at the antenna is

$$P_{\text{RF,out}}(\theta, \phi) = \frac{P_T}{4\pi r^2} G_T(\theta, \phi) A_{\text{eff}}(\theta, \phi) [1 - |\Gamma_{\text{in}}|^2] G_{\text{amp}}$$

where generally the transmitted power P_T and transmitter (test) antenna gain G_T are known, while the effective area of the active antenna, match to the amplifier, and gain of the amplifier are not known exactly. By measuring the RF output power over the full sphere for a given polarization at a given frequency, the quantity $A_{\text{eff}}(\theta, \phi) [1 - |\Gamma_{\text{in}}|^2] G_{\text{amp}}$ is determined. Assuming that for the given value of amplifier gain, the antenna is matched to the input of the amplifier, the above can be expressed as

$$P_{\text{RF,out}}(\theta, \phi) = \frac{P_T}{4\pi r^2} G_T(\theta, \phi) \frac{\lambda^2 G_{\text{ant}}(\theta, \phi)}{4\pi} G_{\text{amp}} = \left(\frac{\lambda}{4\pi r} \right)^2 P_T G_T(\theta, \phi) \cdot G_{AA}(\theta, \phi)$$

where G_{AA} is the gain of the active antenna, which will give $A_{\text{eff}} > A_{\text{geom}}$, or an aperture efficiency greater than 100 percent. In this definition, the DC power to the amplifier is not taken into account, and the efficiency of the antenna, which includes aperture and amplifier efficiency, can be found from

$$\eta_{AA} = \eta_{\text{ant}} \eta_{\text{amp}} = \frac{A_{AA}}{A_{\text{geom}}} \frac{P_{\text{RFout}}}{P_{\text{DC}}}$$

where $A_{AA} = G_{AA} \lambda^2 / (4\pi)$ is the effective area of the active antenna.

32.2 FREQUENCY-AGILE ACTIVE ANTENNAS

The feed impedance and radiation pattern of a frequency-agile active antenna is tuned within a frequency range or in multiple bands. If the electrical properties of the antenna are adjusted in a continuous manner, the antenna is said to be *tunable*, and if the variations are in discrete steps, the antenna is said to be *reconfigurable*. Varactor diodes, photo diodes, p-i-n diodes, MEMS switches, and ferrites are examples of active devices that can be integrated with an antenna in such a way that the antenna is impedance matched at different frequencies for different states of the active device. When the impedance changes, it is usually desired that the radiation pattern, polarization, isolation, efficiency, and gain remain the same. However, integration of devices that necessarily have insertion loss results in deterioration in the antenna efficiency.²⁸ This is what makes the design of frequency-agile active antennas challenging. Note that reconfigurability can also be applied to the radiation pattern or polarization of an antenna, as overviewed in more detail in Chapter 31.

Design and Characterization of Frequency-Agile Antennas

The design of a continuously tunable (e.g. Al-Charchafchi and Frances²⁹) or multiple-frequency-band (e.g., Ollikainen et al³⁰) antenna depends on the desired bandwidth for return loss, polarization, radiation pattern, and gain (efficiency). For example, if switching between two or more bands is required to cover an up- and downlink of a half-duplex communication link,³⁰ in addition to requiring the same return loss at the E-GSM band frequencies of 880–915 MHz and 925–960 MHz, the design goals are minimization of losses in the tuning circuit, similar radiation efficiencies, and similar patterns in the two bands. A sketch of the antenna from Ollikainen³⁰ is shown in Figure 32-6a, while the measured dual-band return loss is shown in Figure 32-6b.

The *antenna element design* is based on bandwidth, polarization, available real estate, and preferred circuit architecture. For continuously tunable antennas, a broadband antenna element is desired. For example, a tapered-slot Vivaldi antenna, which can typically achieve octave bandwidth, has been shown to have a 14% varactor-tuned transmit bandwidth,³¹ defined as a 0.8-dB variation in output power. A second-resonant slot antenna with a possible 15–20% bandwidth was tuned over a 10% bandwidth in Forman and Popovic,⁴ as shown in Figure 32-7. Other examples include microstrip antennas and folded-slot antennas, dipoles, etc. For frequency reconfigurability, the antenna itself does not need to be broadband. For example, resonant slots³² and dipoles³³ have been demonstrated with four-band and three-band operation, respectively. Envisioned applications are frequency-hopped spread spectrum systems where a reconfigurable narrowband antenna follows a pseudo-random pattern of the frequency-hopped modulation. Fractal antennas integrated with MEMS switches have been demonstrated as well.³⁴

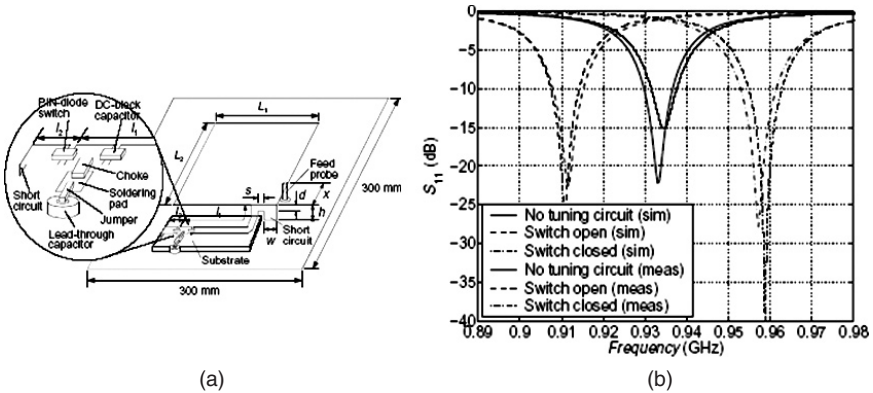


FIGURE 32-6 (a) Schematic of a patch antenna with integrated pin-diode switch and (b) response in the two E-GSM bands. The radiation efficiency was comparable in the two bands, and the radiation patterns remained the same (J. Ollikainen et al.³⁰ © IEEE 2002).

The type of *active element and circuit architecture* are dictated by frequency of operation, and antenna feed. The most commonly used active solid-state devices in frequency-agile antennas are varactor and p-i-n diodes, for continuous and switched operation, respectively. There have also been demonstrations with photo diodes used instead of a varactor, such that the impedance changes as the optical power is varied. In addition, MEMS switches have been proposed as low-loss switching elements. Another possibility is the use of a material with DC bias tuned properties, such as in Castro-Vilaro and Rodriguez-Solis³⁵ where a thin-film ferroelectric material, barium strontium titanate (BST), is used to produce a tunable folded-slot antenna. The DC biasing voltage varies the thin-film dielectric constant, causing a resonant frequency shift. The permittivity of the ferroelectric material can change from 400 to 1200, depending on biasing conditions. In this range the resonant frequency changes from 33.54 GHz to 31.35 GHz, giving an 11% frequency bandwidth with VSWR < 2.

Characterization of frequency-agile antennas should include measurement of return loss, radiation pattern, polarization properties, and efficiency. For example, as can be seen

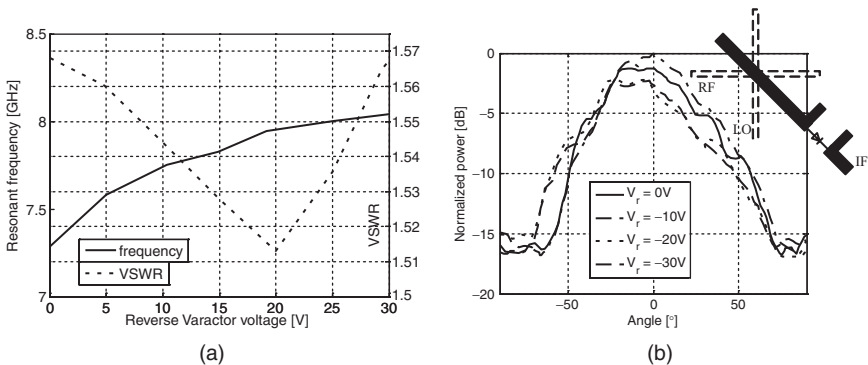


FIGURE 32-7 Second-resonant slot antenna tuned over a 20% bandwidth with a varactor-loaded microstrip feedline (after M. Forman and Z. Popovic⁴ © IEEE 1997). Measured tuning range and corresponding VSWR (a). The measured radiation patterns change as the bias (frequency) is varied (b).

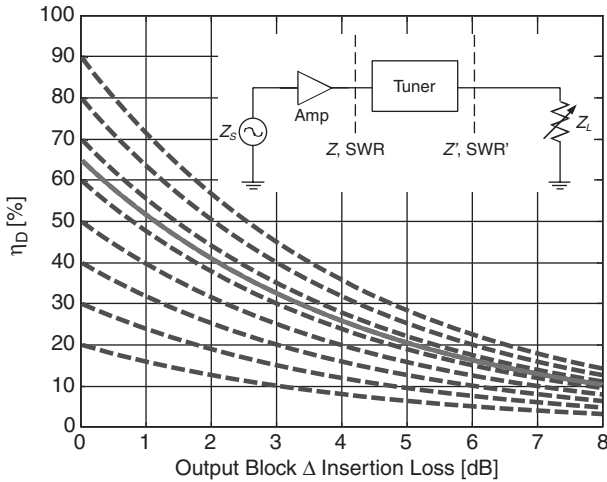


FIGURE 32-8 An amplifier matched to a tuned antenna with a realistic lossy tuner circuit. The drain efficiency is plotted as a function of the insertion loss in the output network for amplifiers with efficiencies from 20 to 90%. The solid black line indicates how the η_D of a 10-GHz class-E PA with 65% η_D would decrease as loss in the output network increases.

in Figure 32-7b, the radiation pattern of the antenna changes over the tuning bandwidth. In this case, the varactor diode cathode is DC- and RF-short to the ground plane, and an extra $\lambda/4$ section is added at the end of the feedline to preserve the RF open. By varying the reverse voltage of the varactor diode from 0 to 30 V, the antenna can be tuned from 7.3 to 8.1 GHz with a VSWR below 1.57, and the peak directivity varies by 2.5 dB over the tuning range. The loss and complexity added by the tuning circuit may not be justifiable. As an example, consider a transmitting active antenna where the goal of the tuning is to provide an impedance match to the amplifier output at several frequencies in a 50- Ω system (see Figure 32-8). Power is lost if either the amplifier is mismatched or there are losses in the circuit/antenna. Assuming a fairly extreme case of a high-efficiency amplifier with 65% efficiency, the question is how much tuning-circuit loss can be tolerated to justify the added complexity. Figure 32-8 shows the calculated efficiency drop as a function of tuner loss. This means that for this high-efficiency case, a tuner/antenna insertion loss of at most 0.3 dB can be tolerated. If the loss is higher, it is more reasonable to tolerate mismatch at the output of the amplifier.

32.3 OSCILLATOR ANTENNAS AND ARRAYS

Antennas integrated with oscillators for power combining and sensors/receivers have been overviewed by a number of authors.³⁶ When an antenna is closely integrated with an oscillator, there is no direct access to the antenna feed or the oscillator output. The design of the active antenna requires both circuit and field modeling. The antenna is part of the feedback loop in an oscillator antenna. It is useful to start the design by considering the antenna to be a simple resonant circuit, as shown in Figure 32-9c,²⁷ although this is a

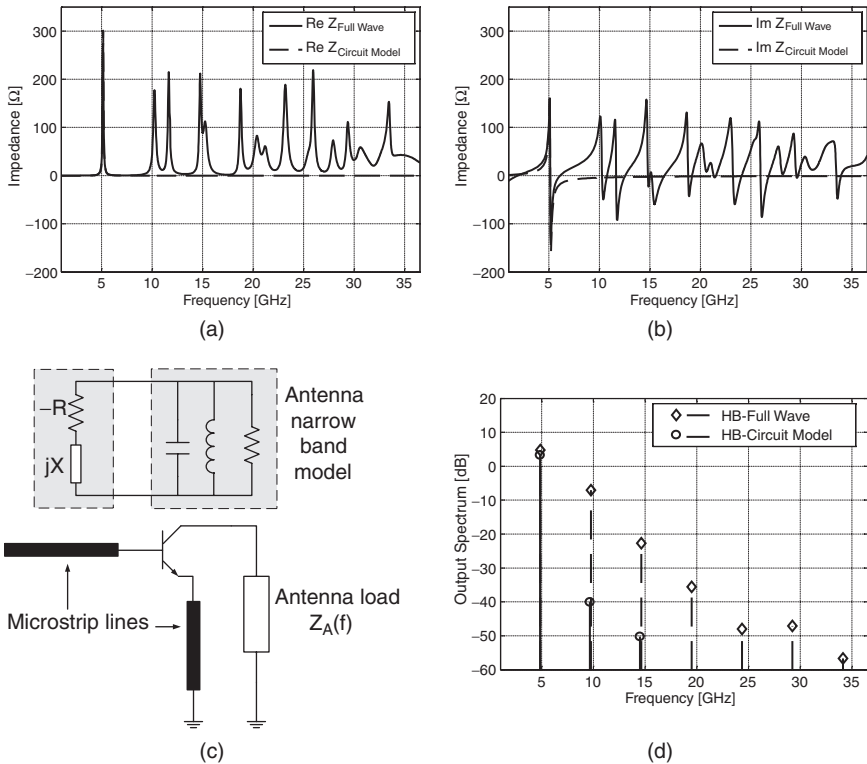


FIGURE 32-9 Example of simulated impedance of a patch antenna over a broad frequency range: (a) real and (b) imaginary part. (c) Simple resonant circuit model of oscillator antenna around first resonance and more realistic model of transistor oscillator antenna. (d) Harmonic balance analysis for the two antenna models, showing the importance of proper broadband modeling.

simplification valid only over a narrow bandwidth. The oscillator circuit is represented as a negative resistance with some equivalent reactance, with bias circuitry omitted. As in standard one-port oscillator models, the negative resistance needs to exceed the radiation resistance of the antenna. In this simple model, the effect of antenna behavior over a broader frequency range is not modeled accurately. Thus, harmonic levels cannot be predicted. An example of the real and imaginary parts of the broadband antenna impedance for the case of an air-patch linearly polarized antenna is shown in Figure 32-9a and b. Both full-wave and circuit models for a patch are included in the plot, showing that the first resonance is correctly modeled by the simple resonant circuit, while the higher harmonics are not. If an active oscillator antenna design is to be correct at the harmonics, the broadband impedance needs to be used in the harmonic-balance nonlinear simulations. A simplified oscillator-antenna circuit is shown in Figure 32-9c. Figure 32-9d compares the harmonic levels obtained with a harmonic balance nonlinear simulation for two circuits which differ only in the way the antenna is included: in one case as the resonant-circuit model and in the other as a set of S -parameters obtained from the full-wave simulation. Power levels at all the harmonics differ considerably in the two cases, including the power at the fundamental. This simple example illustrates the importance of correctly modeling the antenna in circuit design.

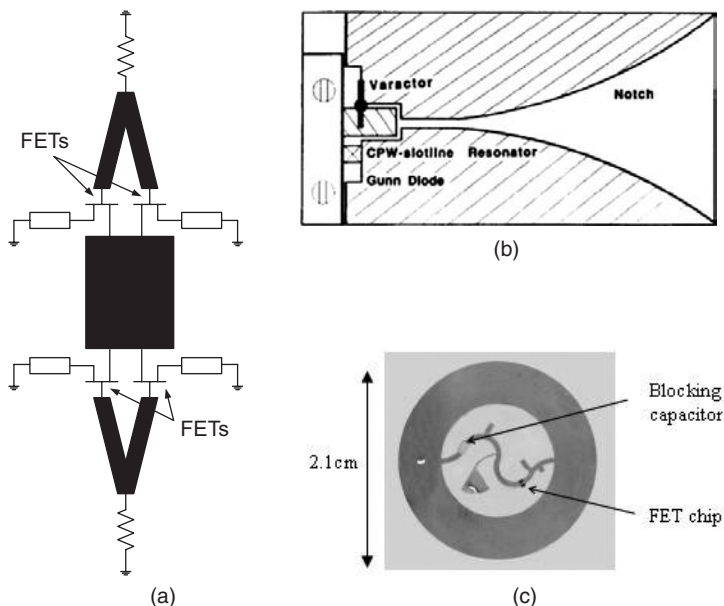


FIGURE 32-10 Examples of oscillator antennas: (a) a push-pull oscillator patch; (b) a Vivaldi slot VCO antenna with varactor-tuned Gunn-diode circuit; and (c) a microstrip annular ring is both the radiating element and microstrip feedback circuit for a class-E amplifier.

For a narrowband transmitter, a narrowband antenna is chosen. If radiation in the half plane is desired, a patch antenna is easily integrated with the oscillator circuit, as in the example in Figure 32-10a.¹⁴ The patch antenna has several other properties that can be advantageous for active antennas: the voltage null in the middle is a natural DC bias line, and the phasing of the voltage and current on the two halves of the patch nonradiating edge allows for multitransistor circuit designs, such as in Figure 32-10a where a push-pull FET oscillator loads a 6-GHz patch antenna giving an output ERP of around 18 dBm. Several such elements have been successfully combined in small arrays with increased ERPs.

For a broadband medium-power tunable linearly polarized transmit antenna element, the Gunn-diode varactor-tuned Vivaldi antenna is chosen in Navarro et al.³¹ (see Figure 32-10b). For millimeter-wave (155 and 252 GHz) low-power transmitters, slot antennas with high cutoff frequency transistors (HFETs) are fabricated monolithically. If unidirectional radiation is required, the slot can be backed by a ground plane, polarizer, or substrate lens. Fundamental oscillations from 115-GHz and 215-GHz monolithic oscillator antennas are reported with 10uW and 1uW estimated output powers from the oscillators.³⁷

Classical design of oscillators based on an amplifier with appropriate feedback can also be applied to oscillator antennas. In a high-efficiency 10-GHz integrated transmitting antenna,³⁸ an annular ring is used both as the radiating element and microstrip feedback circuit for a class-E amplifier. A maximum conversion efficiency of the DC power consumption to radiated copolarized power is 55% at 10 GHz with maximum ERP of 23.6 dBm and total radiated power of 15.5 dBm. In this case, the size of the active antenna is minimal, since the antenna is used directly as the feedback circuit, with an amplifier embedded within the antenna footprint. The annular microstrip ring antenna is operated in its TM_{12} mode. The wider bandwidth of the TM_{12} mode compared to the TM_{11} fundamental mode is

useful in the active antenna design, since it enables some tolerances on the active circuit. Nonlinear models of active devices have difficulty predicting behavior in switched mode of operation, which is compensated for in design by increased impedance bandwidth.

Both two-port and three-port active devices have been used in oscillator antennas: Gunn diodes,³¹ IMPATTs,³⁹ various types of transistors, tunnel diodes,³⁶ and Josephson junctions.⁴⁰ Transistor transmitting antennas currently dominate because of the relative ease of integration, the fact that frequency and power are dictated mostly by the circuit/antenna, and availability of devices that cover a broad range of frequencies and powers.

Application Example for a Single Oscillator Element and Small Array

Single oscillating antennas have been developed for compact low-cost RF sensors, while oscillator arrays have been applied to spatial power combining and phase-shifterless beamsteering. Low-cost, low-power manufacturable automotive crash sensors and RFID tag systems⁴¹ have been developed at 24 GHz for use in the ISM unlicensed band. An example of a sensor antenna is shown in Figure 32-11*a*. The active device at 24 GHz is an HEMT or HBT, and the circuit is implemented in microstrip on an Alumina substrate ($\epsilon_r = 9.8$) with gold metalization and is packaged in a standard package. This active antenna is designed as a three-port oscillator, where the gate microstrip resonator is the radiating element.⁴²

In low-power sensor applications, the parameters of interest are the precise oscillation frequency, frequency stability, and phase noise. We first consider design of the oscillator antenna such that for every device, irrespective of device variations, operation within the 200-MHz band around 24 GHz is ensured. The first step is to accurately model the embedding impedances at the active device terminals with full-wave analysis. This is followed by either a linearized or nonlinear circuit analysis. In the former, the oscillator is considered as a combination of an active multiport and a passive multiport which is obtained from field simulations. The oscillation condition can be expressed as

$$|\det(\mathbf{S}_T \cdot \mathbf{S}_p - \mathbf{I})| > 0 \quad \text{and} \quad \text{Arg det}(\mathbf{S}_T \cdot \mathbf{S}_p - \mathbf{I}) = 0$$

where \mathbf{I} is the identity matrix. This condition is plotted as a function of frequency in the complex plane and crossing of the positive real x -axis at the desired oscillation frequency (24 GHz in this case) indicates an oscillation. It is good practice to measure the S -parameters of the device for varying input power levels. The major effect under large signal conditions is the change in $|S_{21}|$, and a reduction in $|S_{21}|$ by a factor between 0.4 and 0.6 is usually appropriate. This type of analysis usually predicts the oscillation frequency within 1% to 5%

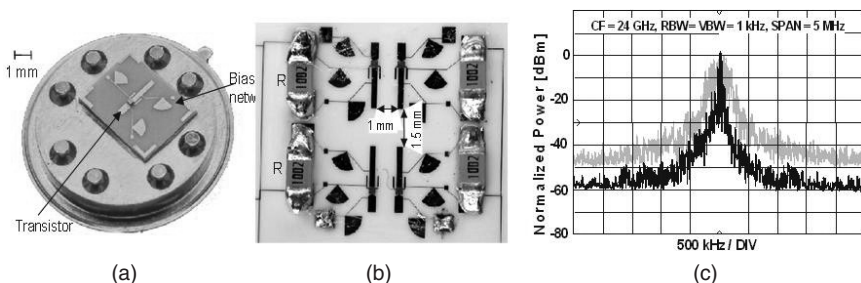


FIGURE 32-11 (a) Photograph of the 24-GHz packaged oscillator antenna using an HEMT device; (b) photograph of four-element active array, and (c) measured spectrum of array (black) compared to free-running spectrum of single element (gray)

of the measured value, but does not give any prediction of power or harmonic levels. To obtain the exact oscillation frequency, a varactor diode weakly inductively coupled to the gate resonator can provide 3% tuning.

An important parameter for sensors that rely on small frequency modulations (e.g., Doppler shifts) is the phase noise. The phase noise can be improved by increasing the loaded Q factor of the resonator antenna, by injection locking, or by using mutual locking of a number of active antennas in a small array, and is inversely proportional to the square of the loaded Q factor of the resonator-radiator. When a longer gate resonator-radiator with doubled unloaded Q factor is inserted in the same active antenna, a decrease of 5 dB, compared to the expected 6 dB, in phase noise is measured. The different resonator will, however, affect the radiation pattern and polarization. Injection of a low-phase noise external source can reduce the phase noise further (down to -80 dBc/Hz at 100-kHz offset in this case) but this approach is not practical for low-cost small sensors because it requires a secondary source. In an array of self-locked oscillators, the signal combines coherently and the noise incoherently, resulting in improved phase noise. Mutual locking of a number of nonlinear oscillators cannot be simulated with commercial software tools. However, coupling in the passive part of the circuit can be simulated and used as a useful starting point for mutually locked array design. The coupling for a four-element array shown in Figure 32-11b can be described by a symmetric 4×4 admittance matrix. The oscillation condition can then be calculated by

$$-Y_{\text{osc},i}(\omega_r, A_i) = \sum_{j=1}^N Y_{ij}(\omega_r)$$

where $Y_{\text{osc},i}$ is the frequency and amplitude-dependent transistor admittance, and Y_{ij} are the elements of the coupling matrix. Based on simulation, a four-element array was implemented, with phase noise compared to a single oscillator antenna, as shown in Figure 32-11c. The ERP increased by roughly 10 dB, while the phase noise decreased by 7 dB. Both numbers differ from the theoretical maximum when four identical elements are locked coherently without change in operating conditions. In the array, however, due to the coupling, the individual oscillators operate under different conditions than when alone, and their biases need to be adjusted for locking at 24 GHz. The expected 12-dB increase in ERP (6 dB in power and 6 dB in directivity) and 6-dB decrease in phase noise can only be design guidelines.

Oscillator Antenna Arrays

In addition to phase noise reduction, arrays of oscillator antennas have been implemented for power combining of low-power oscillators. In arrays with standard half-wavelength spacing, the coupling between antenna elements may not be sufficient to enable oscillator synchronization through injection locking, and additional coupling needs to be provided, e.g. through a circuit. This approach lends itself to beam-steering through frequency control. Alternatively, grid arrays with a period on the order of $\lambda/10$ are inherently tightly coupled and provide the highest power densities in power-combining active apertures.²⁷ Quantities that are directly measurable are ERP, radiation pattern, polarization properties, spectral content, and input DC power. From these quantities, the radiated RF power and conversion efficiency can be estimated.

Most standard arrays with half-wavelength spacing between elements have been limited in the number of elements to 16, due to the difficulty in maintaining synchronized oscillations. In tightly coupled grid oscillators, however, the number of elements that can be coherently combined is in the hundreds. The first such planar grid oscillator with 100 MESFET elements at about $\lambda/10$ spacing at 5 GHz is shown in Figure 32-12a,⁸ where λ is the free-space wavelength. The vertical metal lines are connected to the drain and gate leads, and the

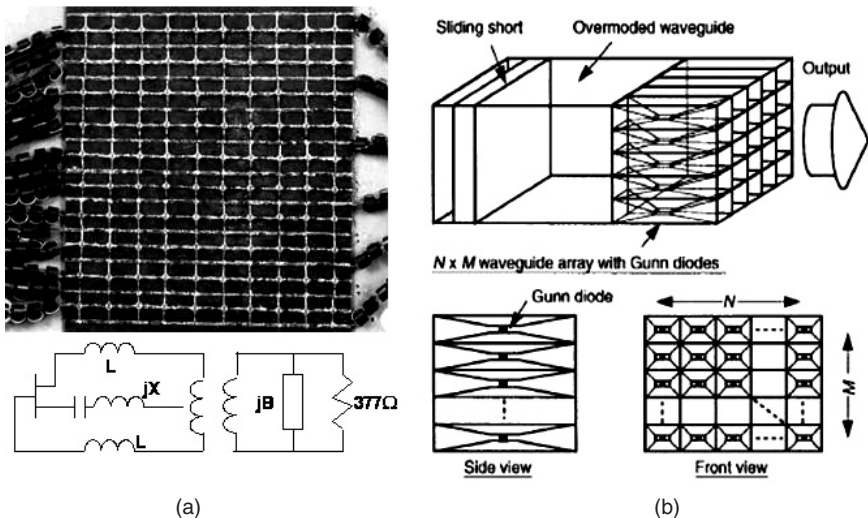


FIGURE 32-12 (a) 100-MESFET grid oscillator with measured ERP = 22W (after Z. Popovic et al⁸ © IEEE 1991). The grid is linearly polarized with a cross-polarization level of -30 dB at broadside. An equivalent circuit convenient for grid-oscillator analysis is shown. (b) Overmoded waveguide resonator coupled to array of waveguides combines a 2D array of Gunn-diode oscillators at millimeter-wave frequencies (after Bae et al⁴⁴ © IEEE 1998).

horizontal metal lines to the source leads and V_{GS} and V_{DS} bias supplies. The grid is a square around λ on the side, backed by a flat reflector of the same size. When the transistors are biased, oscillations start from noise. Power is radiated incoherently from the drain terminals, reflected off the mirror, and received by the gate terminals. For appropriate mirror spacing combined with the phase between gate and drain voltages, steady-state locked oscillation is reached through global coupling after several bounces inside the resonator.

Analysis of this combiner is based on the assumption that the symmetry of the grid reduces it to a unit cell with electric walls on the top and bottom and magnetic walls on the sides. The passive multiport connected to the transistor terminals can be found using the EMF method⁸ or commercial software such as Ansoft's HFSS. An equivalent circuit that describes the physical behavior of the grid is shown in Figure 32-12b, where the vertical leads are represented by inductances, the horizontal leads by a resonant circuit, and the currents on the leads are coupled to the plane wave (a transmission line with a 377- Ω characteristic impedance) through a center-tapped transformer. The mirror is represented by a short circuit. Circuit simulations using linear parameters give a resonant frequency within 1 to 5% of the measured ones.

Subsequent grid oscillators have demonstrated operation at millimeter-wave frequencies and with up to 660W ERP and 10W of radiated power from 100 devices at X-band.⁴³ Similar arrays at millimeter-wave frequencies where the devices are embedded in one of the mirrors of the resonator are described in, e.g., Bae et al,⁴⁴ where an overmoded-waveguide resonator with an array of TE₁₀-mode waveguides containing Gunn diodes achieved highly efficient spatial power combining at V-band (61.4 GHz). An output power of 1.5W CW was measured with an estimated efficiency of 83% with a 3×3 Gunn-diode array, Figure 32-12c.

Large-scale power combining in a grid oscillator is achieved after radiation in free space. For all elements in phase, the power is combined in the broadside direction and the grid radiates like a continuous active aperture producing a radiation pattern close to $(\sin x/x)^2$. For applications where beam-steering is of interest, antenna arrays with half-wave spacing

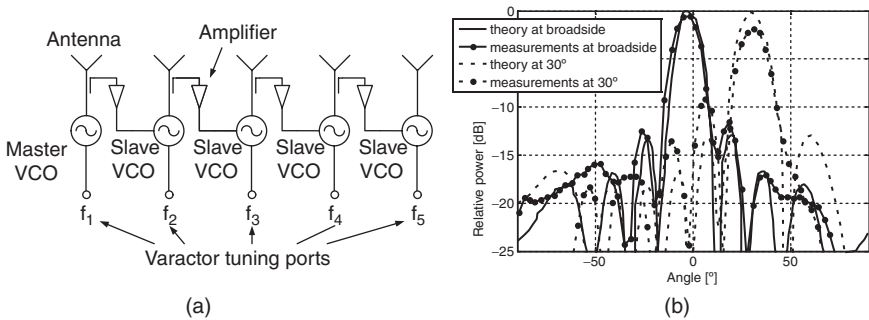


FIGURE 32-13 (a) Principle of beam-steering by frequency detuning. (b) Example measured radiation patterns as the control voltages of the edge VCOs are varied to produce a frequency difference across the array, corresponding to a phase gradient (after R. A. York and T. Itoh⁴⁵ © IEEE 1998).

are more appropriate. A very interesting and elegant approach is described in detail in Popovic and York²⁷ and is illustrated in Figure 32-13. The oscillator antenna elements at the edge of the linear array are detuned in frequency, while the middle elements lock with progressive phase, corresponding to a differential frequency between the end elements. In an X-band prototype ten-element power-combining beam-scanning array of coupled voltage-controlled oscillators (VCOs)⁴⁵ and varactor-tuned patch antennas, a measured ERP of 10.5W at 8.4 GHz is achieved with a maximum scan angle of 30 degrees, corresponding to 500 degrees of phase shift across the array. A theoretical investigation of the phase noise in mutually synchronized oscillator systems indicates the necessity of using external injection locking. The steady-state phases in coupled oscillator arrays are dependent not only on the free-running frequency distribution, but also on the coupling-phase angle, as shown in the insightful theory by York.⁴⁵ Beam-steering arrays with PLLs and multipliers have also been demonstrated. However, extensions to 2D arrays seem unlikely, because it is difficult to independently control and design coupling in two dimensions.

32.4 AMPLIFIER ANTENNAS AND ARRAYS

Amplifier antennas fall into one of the three categories described in Figure 32-1c: receiving antenna integrated with LNA, transmitter PA with antenna load, and amplifier with antenna input and output. While the first two cases are useful both as single elements and arrays, the latter is used mainly in arrays for spatial power combining. The antenna gain is increased compared to a passive antenna at the expense of input DC and RF power. The normalized antenna radiation pattern should remain the same, although addition of circuit and bias can change the antenna far-field properties, especially polarization and sidelobes. The bandwidth of the antenna tends to increase, since effectively loss is added and the Q factor is reduced compared to the passive antenna since the amplifier efficiency is always lower than 100 percent.

Single Amplifier Antenna Elements: Receiving and Transmitting

In a receiver, any loss in the antenna and feedline connected to the LNA contributes to the reduction of noise figure as $F_{\text{tot}} = F_{\text{feed}} + L(F_1 - 1)$, where F_1 is the noise figure of the first amplifier in the receiver chain and L is the loss of the antenna feed connected to the input of the LNA. Usually, the optimal noise match at the input of the amplifier is not 50 Ω ,

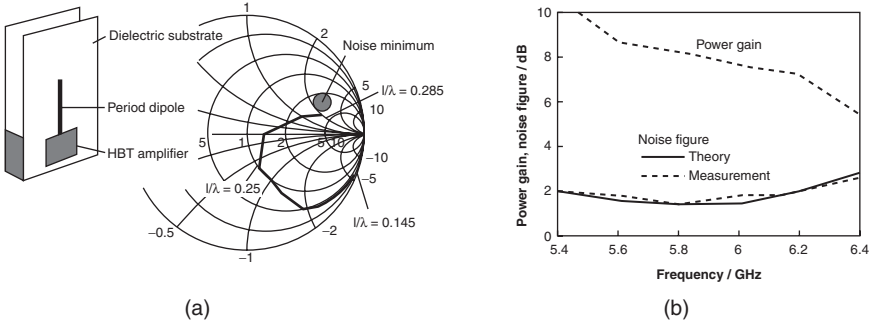


FIGURE 32-14 (a) Input reflection coefficient of printed monopole antenna as a function of length plotted along with the optimal noise match for a SiGe HBT device. (b) Measured and calculated noise figure show increased bandwidth for the integrated amplifier antenna. The power gain is relative to a passive antenna (W. Duerr et al.⁴⁶ © IEEE 1997).

and is also not simultaneously an optimal input reflection coefficient match. In Duerr et al.,⁴⁶ the impedance of a monopole antenna for different dipole lengths was examined as a direct noise match for a SiGe HBT amplifier in the 5.8-GHz WLAN band. Figure 32-14 shows the relevant antenna and device impedances, as well as a plot of the noise figure.

In Radulovic et al.,⁴⁷ a patch antenna is designed to present complex optimal noise figure impedance to a HEMT device at 13 GHz. The patch is fed with an indented microstrip line, which allows for design of complex input impedance by varying two geometric parameters for a given substrate. It is shown that when the patch antenna is designed to provide an optimal noise reflection coefficient of $0.28 \angle 150^\circ$, the radiation pattern of the antenna remains the same, but the gain increases from 7.7 dBi for the passive antenna to 15.1 dBi for the active amplifier antenna. It is interesting to note that there have been demonstrations of broadband electrically short half-loop antenna amplifiers (50–500 MHz),⁴⁸ which are cryogenically cooled and integrated with a SQUID array and buffer amplifier in a portable package. The addition of the inductive SQUID array and buffer amplifier provided the broadband matching for the small inductive antenna.

It is not straightforward to measure the noise figure of the amplifier in an active receiving antenna. If the antenna temperature is T_A , the noise figure of the amplifier can be found from

$$F = 1 + \frac{N}{G_{AA}} - \frac{T_A}{290}$$

where N is the noise spectral power density measured at the output of the LNA referred to input at room temperature of 290K (−174 dBm/Hz), and G_{AA} is the transducer gain of the active antenna, which in turn requires measurements of a reference passive antenna.⁴⁹ The ambient temperature (T_A) needs to be well controlled.

In a transmitter, the antenna is the output load to the amplifier. For a power amplifier, the optimal transistor output impedance is usually not 50 Ω , but is determined by the mode in which the transistor operates. For example, in the case of a class-A saturated power amplifier (PA), the load line and device and packaging parasitics determine the optimal output power impedance. In the case of a high-efficiency switched-mode PA in class-E mode of operation, a specific output impedance forces the time-domain current and voltage waveform product, $i(t)v(t)$, to be zero during most of the period of the carrier frequency, thus minimizing the loss in the transistor. For soft switching, it can be shown that this complex impedance is, e.g., Weiss and Popovic,²⁶

$$Z_E = \frac{0.28015}{\omega C_{\text{out}}} e^{j49.052^\circ}$$

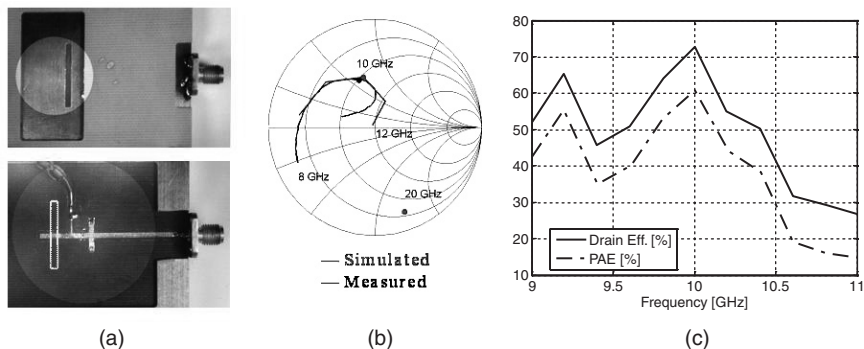


FIGURE 32-15 (a) Circuit side and antenna side of active class-E amplifier antenna at 10 GHz. (b) Simulated and measured nonresonant slot antenna impedance, along with optimal class-E impedances for transistor at fundamental (10 GHz) and second harmonic (20 GHz). (c) Measured efficiency of active antenna as a function of frequency.

where $\omega = 2\pi f$ is the fundamental operating angular frequency, and C_{out} is the output capacitance of the transistor. The higher harmonics are assumed to be terminated with high impedances (ideally open circuits). In Weiss and Popovic,²⁶ a 10-GHz active class-E antenna is demonstrated with a MESFET device and a slot antenna (see Figure 32-15a). The output capacitance of the transistor is 0.107pF, resulting in an optimal output impedance of $41.7\Omega \angle 49.05^\circ$ with a resulting reflection coefficient with respect to 50Ω equal to $-6.67\text{dB} \angle 104^\circ$. Thus, a nonresonant antenna with this impedance at the fundamental 10-GHz frequency and a high impedance at the higher harmonic frequencies can be directly connected to the transistor and thus reduce the footprint of the active antenna. A slot antenna is designed using a full-wave simulator, with impedance at the fundamental and second harmonic as shown in Figure 32-15b. A passive slot was also fabricated and the measured reflection coefficient at 10 GHz, with a TRL calibration, was measured to be $-6.8\text{dB} \angle 106^\circ$. The efficiency and power of the antenna are estimated from measured copolar and cross-polar radiation patterns. The radiated power of around 100 mW with 75% drain efficiency was achieved. This is higher or comparable to an amplifier designed with the same device in a microstrip circuit, since some circuit losses are avoided. Due to the direct circuit-antenna integration, the total active antenna footprint is $0.4\lambda^2$.

Receiving Amplifier Antenna Arrays

Figure 32-16 shows an active receiving antenna array where antennas are directly connected to LNAs in each element with no lossy isolators or circulators. Such antenna arrays are used in radioastronomy, either as large-area arrays or reflector feeds. The transistor amplifier produces noise waves at both of its ports, and the noise at the input will radiate, potentially coupling into the neighboring antenna elements and amplified by their LNAs thus degrading the system noise performance.⁵⁰ As the phases of the phase shifters are varied and the beam is scanned, the amount of coupling between the antennas varies and this affects the reflection coefficient at the input of each of the LNAs, as well as the amount of noise coupled to neighboring elements.

Referring to Figure 32-16a, it can be shown that the coupled noise at the output of all the phase shifters is

$$c_{\text{tot}}(\theta, \phi) = c_2 + Gc_1S_i(\theta, \phi)$$

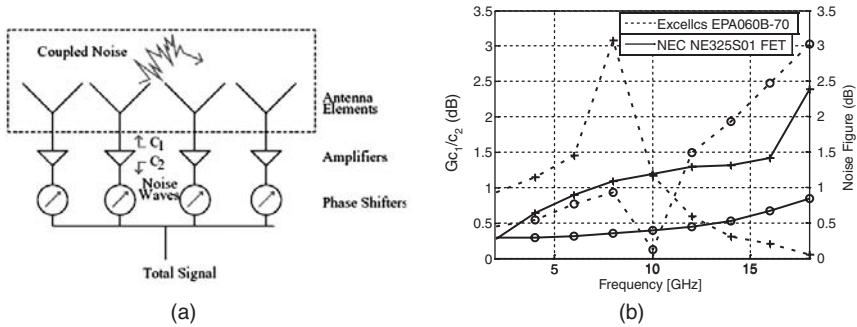


FIGURE 32-16 (a) Receiving LNA active array with noise coupling. (b) Plot of transistor characteristics relevant to noise coupling. The best noise figure of a transistor does not correspond to a minimal noise coupling coefficient for a given antenna array.

where c_2, c_1 are the noise waves at the input and output of all (identical) LNAs, G is the gain of each LNA, and $S_i(\theta, \phi)$ is the active scan reflection coefficient of element i for a scan angle (θ, ϕ) . The noise coupling for a given scan angle can therefore be found from measured or simulated active scan reflection coefficient, amplifier gain, and amplifier input and output noise waves. The latter can be determined using standard techniques from minimal noise figure F_{\min} , optimal reflection coefficient Γ_{opt} , and noise resistance R_n , parameters given in specification sheets for most transistors. If a blind (noisy) scan angle appears for a certain element type or element coupling, the noise coupling can be modified by changing the antenna type, as well as the transistor.

As an example, consider a 7×7 half-wave dipole array above a ground plane with a half-wave spacing. The 49×49 full scattering matrix of the array can be found using a full-wave simulator, and the scan reflection coefficient can be found for each element. The noise coupling ratio, defined as the ratio of the uncoupled noise to the coupled noise, can be written as $\text{NCR} = (Gc_1 S_i) / c_2$ and is a function of antenna array properties and amplifier properties. The amplifier quantity Gc_1 / c_2 is plotted for three different transistors in Figure 32-16b, while the noise figures for these amplifiers from their data sheets are shown in Figure 32-16c. A lower noise figure can un-intuitively result in higher noise coupling, reflecting the fact that minimal noise figure implies low noise power at the output, but not necessarily at the input of an LNA.

An interesting figure of merit is the amount of coupling that produces 50 percent more noise power at the array output at certain scan angles, as compared to that expected without noise coupling. To find the minimal amount of inter-element antenna coupling that would give the noise ratio of -3 dB, we set

$$10 \log \frac{|Gc_1 S_i(\theta, \phi)|^2}{|c_2|^2} = -3 \text{ dB}$$

and solve for the allowable active scan coefficient for each array element, given the amplifier properties. For reasonable LNA gain, the practical scan angle for low coupled noise will be limited: low noise coupling requires low values of active scan coefficient, i.e., larger element spacing; but larger element spacing limits the scan angle due to sidelobe appearance. The design of a low-noise scanning array therefore involves choice of transistor, design of LNA and match to antennas, choice of antenna element, and design of array for minimal coupling for desired scan range.

Transmitting Amplifier Antenna Arrays

A power amplifier integrated in each antenna element provides distributed amplification, resulting in higher achievable efficiency and improved degradation properties for a given ERP. Spatial power combining upon radiation results in an ERP that is proportional to N^2 , where N is the number of amplifier antenna elements. Spatially fed combiners are fed in two ways: using a guided-wave feed network and a spatial feed. For a given combiner network loss per stage, there is a minimal number of elements N for which spatial power combining has a larger power-combining efficiency.²⁷

An example of a circuit-fed array is shown in Figure 32-17a where 16 high-efficiency 10-GHz amplifiers are spatially combined using stacked patch antennas spaced a half wavelength apart.⁵¹ In this case, the antennas and feeds are designed to present a specific complex impedance to the output of the transistors.²⁶ A corporate four-level Wilkinson divider feed network is used at the input of the array with a measured loss of 1.4 dB. The output powers are spatially combined through an array of multilayer patch antennas with 2:1 VSWR bandwidth of 11.6%. An average drain efficiency of 70% at 162W EIRP, or approximately 1.5W of transmitted power, is measured for an array of 16 amplifiers at 10.2 GHz. The estimated power-combining efficiency is 75%.

Spatially fed, spatially combined arrays have also been demonstrated, e.g. using slot antennas at Ka-band as shown in Figure 32-18a. Input and output antennas are anti-resonant cross-polarized slots fabricated on alumina⁵² or aluminum-nitride⁵³ substrates with MMIC amplifiers providing power gain between the input and output slots. Polarizers ensure unidirectional slot radiation. Several such arrays with watt-level power at Ka-band have been demonstrated, e.g. a 6×6 array gave an output power of 4W and power gain of 6 dB at 29 GHz, where a liquid-cooling test fixture removed excess heat. Aluminum nitride proved to be an adequate heat conductor for a 1W 6×6 33-GHz array. In this case, all input slot antennas are spatially fed with a dielectric-loaded (“hard”) horn antenna in the near field of the array. The amplitude and phase profile of the horn loaded with the array is not uniform and degrades the power-combining efficiency. Figure 32-18b shows the measured H-plane copolarized radiation patterns of two arrays that differed only in bias line configurations: the bias lines of Array B have extra air bridges and capacitors that short slot-mode radiating RF currents on the bias lines, which results in an improved pattern.

A compact approach for combining a smaller number of elements was done in N-S. Cheng et al.^{54,55} where tapered-slot antennas connected to both ends of 50-Ω monolithic amplifier chips are inserted in a waveguide, achieving broadband combining from 8 to 11 GHz with over 120W of output power. In the array shown in Figure 32-19a, a total of $N = 24$ amplifiers is combined, with six cards inserted in the waveguide, containing four MMIC amplifiers each. This approach provides large bandwidth, good heat-sinking, and space

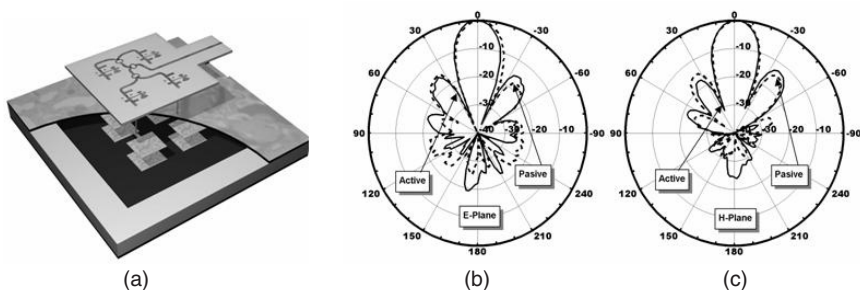


FIGURE 32-17 (a) Schematic of the four-element active subarray showing multiple layers. (b) Radiation patterns of the passive and active 16-element antenna arrays (S. Pajic and Z. Popovic⁵¹ © IEEE 2003).

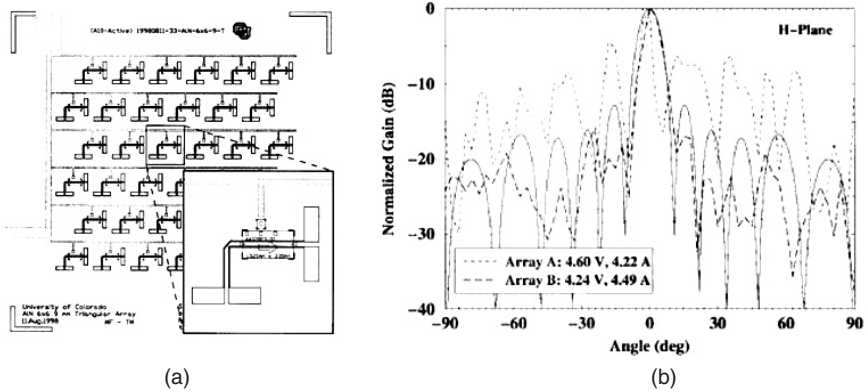


FIGURE 32-18 (a) Layout of a Ka-band slot antenna amplifier array with prematched 50- Ω MMIC amplifiers. Polarizers are used externally to the array for unidirectional slot radiation. (b) Measured H-plane copolarized radiation patterns for two arrays that differ only in bias-line implementation (T. Marshall et al.⁵³ © IEEE 1999).

in the z -direction to add gain, and has graceful degradation as shown in Figure 32-19b. However, the architecture accommodates a limited number of cards, and the waveguide mode has a nonuniform input field profile, thus the edge amplifiers contribute less to the total output power.

Another approach for large-scale combining is a grid amplifier.⁵⁶ The radiating portion of a grid amplifier is similar to that of a grid oscillator (see Figure 32-12), except in this case, the input and output radiating elements are orthogonally polarized to reduce feedback between input and output waves. A differential transistor pair provides amplification in each unit cell. Figure 32-20 shows a photograph of a portion of a monolithic 37-GHz grid amplifier, with measured 5W of output power.⁵⁷ Grid amplifiers have been realized as both transmission and reflection type. In DeLisio et al,⁵⁸ a 31-GHz two-stage transmission

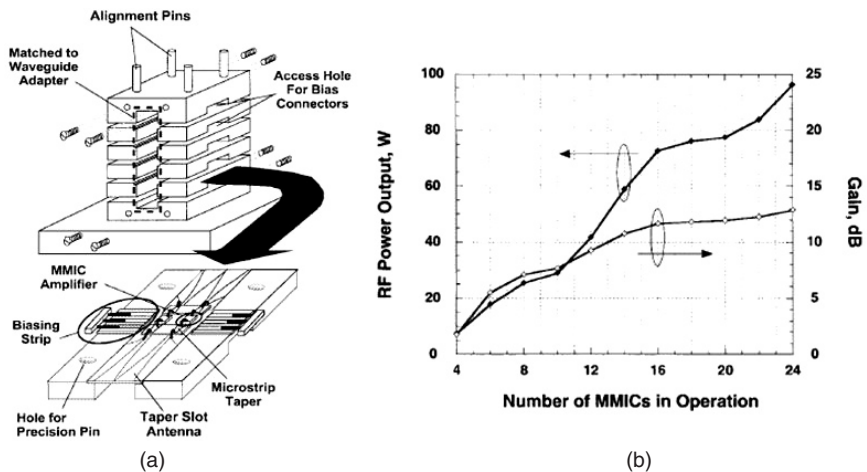


FIGURE 32-19 (a) Architecture of a broadband finline waveguide amplifier array, delivering over 120W over the entire waveguide X-band. (b) Measured gain and output power vs. number of operating amplifiers. The inset shows one card with two tapered slot antennas connected with MMIC amplifiers (N-S. Chang et al.⁵⁵ © IEEE 1992).

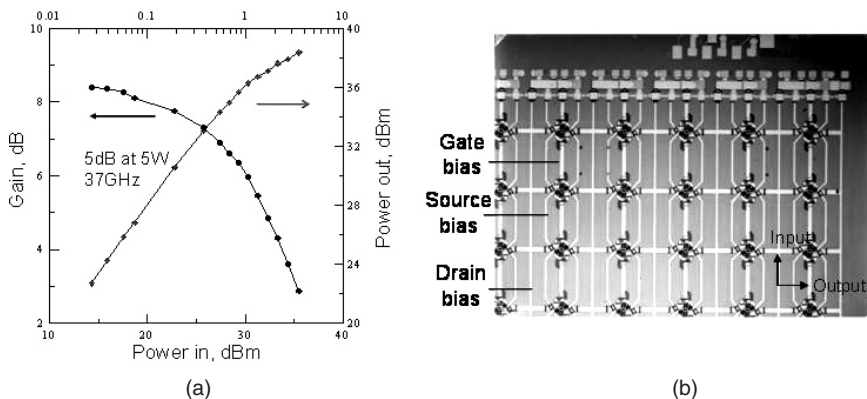


FIGURE 32-20 (a) Measured gain and output power of a 37-GHz monolithically integrated grid amplifier. (b) Photograph of a portion of the 256-element 1-cm-square grid-antenna amplifier chip. The input electric field is vertically polarized, and the amplified wave is radiated in the horizontal polarization. The bias lines are vertical. (Courtesy Prof. D. Rutledge, Caltech)

grid amplifier module with over 10W output power and 12-dB gain is described. An 82-GHz reflection grid amplifier with 5.5-dB gain, 400-MHz 3-dB bandwidth, and 110-mW saturated output power is presented in Cheung.⁵⁹ Commercially available grid amplifiers are not plane-wave fed from the far field, but placed in an overmoded waveguide with excellent feed efficiency, in which case they arguably are no longer active antennas.

Practical issues in active antenna amplifier arrays include feeding, heat management, impedance matching, and stability. Several types of spatial feeds have been demonstrated: plane-wave feeding is inefficient and practical only for characterization; Gaussian-beam feeds are low-loss but large⁵⁸; dielectric-loaded horn antennas do not produce uniform amplitude and phase illumination, thus reducing power-combining efficiency⁶⁰; and discrete lens antenna arrays, e.g. Hollung et al,¹² are large in volume but provide additional functionality. An especially difficult problem is stability in active antennas and arrays—electromagnetic coupling is present due to radiation. Bias-line oscillations are easily excited in amplifiers where one of the loads radiates, and this can limit the gain of the active antenna. Shunt capacitors of different values placed along the bias lines to short the bias-line oscillation currents to ground can eliminate instabilities.

32.5 FREQUENCY-CONVERTING ANTENNAS

The two extreme cases of frequency-converting antennas are oscillator antennas (Section 32.2) that convert DC power to power at a single RF frequency and its harmonics, and rectifier antennas that convert incident RF waves into DC power (Section 32.6). In this section, multiplier and mixer active antennas are briefly overviewed.

Multiplier Antennas and Arrays

Although there have been a number of active antenna frequency multipliers demonstrated in the literature, it is easy to conclude that at microwave frequencies, fundamental-frequency oscillators will give more power and lower phase noise, which grows at best as

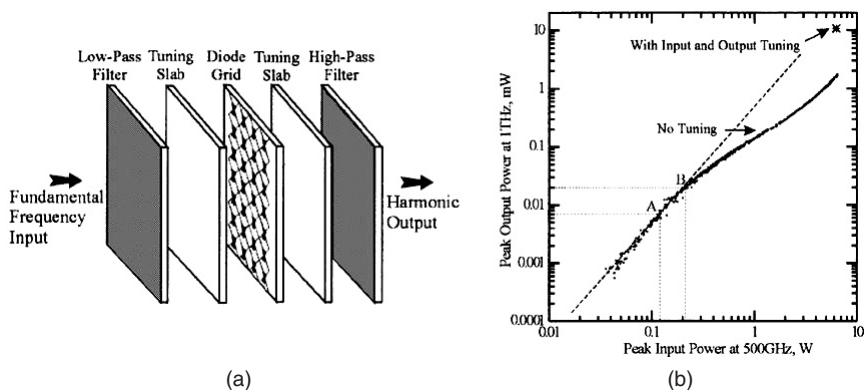


FIGURE 32-21 (a) A millimeter-wave active antenna quasi-optical grid multiplier; (b) measured output power at 1 THz with and without input and output tuning slabs (A. Moussessian et al⁶² © IEEE 1998)

$20 \log N$, where N is the multiplication factor. At millimeter-wave frequencies, however, transistors are either not available or have low output powers, and multiplication becomes more relevant. Frequency doubler and tripler antenna arrays for millimeter-wave generation with Schottky and varactor diodes have the advantage of input power division between a large number of elements, allowing high-power generation. For example, 0.5W was generated at 66 GHz in Jou et al⁶¹ in a monolithic 760-diode grid array with around 10% conversion efficiency from a pulsed source. A similar approach was used by Moussessian et al⁶² where 144 diodes generated 24 mW at 1 THz from a 47W peak pulsed source with an efficiency below 1%. As shown in Figure 32-21, the input plane wave is incident on an array of antennas loaded with a nonlinear harmonic-generating element (diodes). The output wave is filtered so that only the higher harmonic is radiating in transmission. For tripler antennas, the diodes are connected in anti-parallel, eliminating the even harmonics for perfectly matched diode pairs.

Mixer Antennas and Arrays

Among the many interesting up/down-conversion active antennas, reviewed nicely in Mortazawi et al,³⁶ we choose a few examples as illustrations in this chapter. Self-oscillating mixer antennas with fundamental and subharmonic oscillations have been demonstrated, and an example of a 60-GHz balanced integrated-antenna self-oscillating mixer is shown in Figure 32-22.⁶³ This mixer employs the modal radiation characteristics of a dual-feed planar quasi-Yagi antenna to achieve RF–local oscillator (RF–LO) isolation between closely spaced frequencies. The balanced mixer is symmetric, inherently broadband, and does not need an RF balun. A 30-GHz push-pull circuit is used to generate the second harmonic, and a 30-GHz dielectric resonator is used to stabilize the fundamental oscillation frequency. This allows the possibility of building a balanced, low-cost self-contained antenna integrated receiver with low LO leakage for short-range narrowband communication. Figure 32-23b shows the measured conversion efficiency of this mixer antenna at the sum and difference IF ports.

A grid mixer, similar to grid oscillators and amplifiers and attractive for millimeter-wave frequencies, is demonstrated in Hacker et al⁶⁴ at X-band. A planar bow-tie grid periodically loaded with 100 Schottky diodes showed an improvement in dynamic range of 3.5 dB over an equivalent single-diode mixer with the same conversion loss and noise figure. The advantage of a mixer array is that power handling and dynamic range scale as the number of the devices.

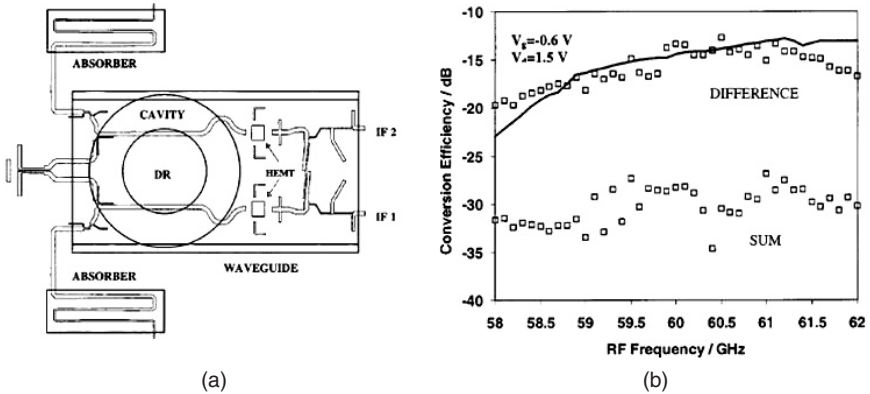


FIGURE 32-22 (a) Layout of a subharmonic mixer active antenna; (b) measured conversion efficiency for sum and difference IF ports as a function of the V-band RF frequency for a 30-GHz push-pull self-oscillating mixer (M. Sironen et al.⁶³ © IEEE 2001)

Different nonlinear functions can be performed by active mixer antennas. An example is four-wave mixing, or phase conjugation, as shown in Pobanz and Itoh,⁶⁵ with the application to saving power in back-scatter mode communication and radar systems. A phase-conjugating active antenna is shown in Figure 32-23a. The circuit is similar to a standard single-balanced mixer with a rat-race hybrid, but with the LO and IF interchanged, so that isolation between RF and IF is accomplished via hybrid balance rather than the usual filter. In this case, the LO frequency is chosen to be twice the RF, and a dual-frequency ring hybrid is implemented. The LO is coupled into the ring through a 12-GHz bandpass filter. The filter is an open circuit at the ring port at 6 GHz, with sum

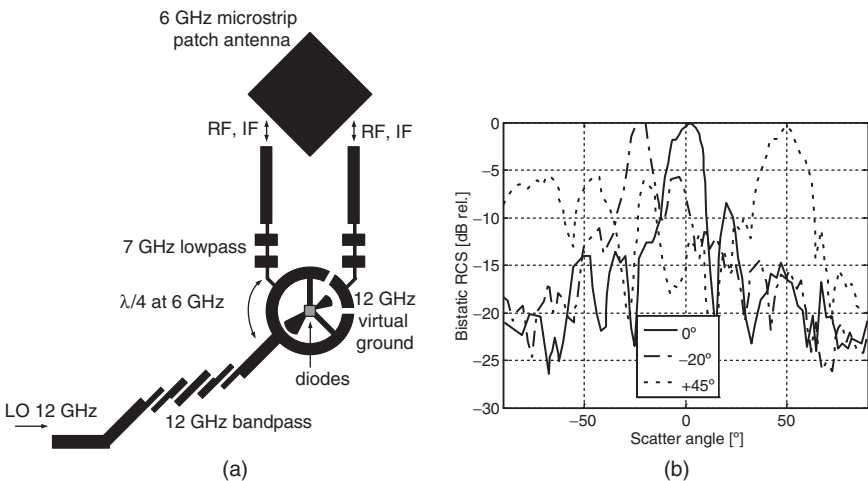


FIGURE 32-23 (a) Layout of a phase-conjugating four-wave mixing active antenna element; (b) measured radiation patterns of an eight-element array for waves incident from broadside, -20° , and 45° (after Pobanz and Itoh⁶⁵ © IEEE 1995)

and difference inputs feeding the mixer diodes at the remaining ports. Low-pass filters (LPFs) are used on the RF/IF ports, which appear as open circuits to the ring at 12 GHz. Since the LO wavelength is half that of the ring design wavelength, only a sum port can be formed to feed the diodes with equal power. By placing the LO port in the longer section of the ring at a 6-GHz quarter-wave from either port 1 or 4, two virtual ground points can be created in the matched ring. The mixer is coupled to an antenna with two orthogonal feeds, allowing this phase-conjugate scattering element to operate for any polarization of the incident wave. Measured radiation patterns for an eight-element array with a 0.8λ spacing with different incidence angles of the incoming wave are shown in Figure 32-23b. In the phase-conjugating process, the spacing between the elements is effectively doubled, resulting in grating lobes even for half-wavelength spacing, which is traded off with mutual coupling between elements.

32.6 RECTENNAS

An integrated antenna and rectifier is usually referred to as a *rectenna*, as shown in Figure 32-24a. The incident waves within a certain spectral range are received by the antenna and coupled to the rectifying device (diode in this case), and the LPF ensures that no RF is input to the power management circuit. A controller provides input to the power management circuit, which enables storage of the received energy over time, and delivery of DC power at the level and time when it is needed. Such active antennas have been primarily proposed for wireless power transmission, which dates back to Nikola Tesla's U.S. Patent No. 685,954 (1901),⁶⁶ which describes wireless transmission of energy, storage of the energy in a capacitor, and energy management in time. Rectification of microwave signals for supplying DC power by high-power beaming has been researched for several decades, and a good review of earlier work is given in Brown.⁶⁷ In power beaming, the antennas have well-defined polarization, and high rectification efficiency is enabled by single-frequency high microwave power densities incident on an array of antennas and rectifying circuits. Applications for this type of power transfer have been proposed for helicopter powering, solar-powered satellite-to-ground power transmission, inter-satellite power transmission including utility power satellites, mechanical actuators for space-based telescopes, small DC motor driving, and short-range wireless power transfer.⁶⁷⁻⁷³ Linear, dual-linear, and circular polarization of the receiving antennas were used for demonstrations of efficiencies ranging from around 85–90% at lower microwave frequencies to around 60% at X-band and around 40% at Ka-band.

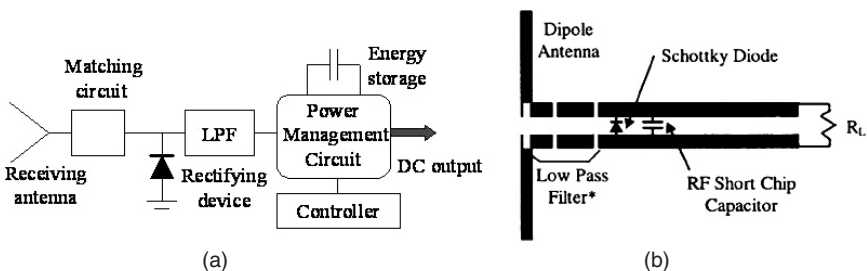


FIGURE 32-24 (a) Schematic of a rectenna and associated power management circuit; (b) an example of a dipole rectenna for 5.8-GHz narrowband operation (after J. O. McSpadden⁷² © IEEE 1998)

There are several distinct scenarios for wireless powering, which influence the rectenna design:

- One or more high-directivity narrowband (single frequency) line-of-sight transmitters with well-known and fixed polarization and well-known power levels. In this approach, used for power beaming, the efficiency of the rectenna can be very high.
- One or more medium-power semidirectional transmitters that illuminate a range in space, with multipath present. In this case, the incident power density is known approximately, but there are multipath effects that change polarization and spatial distribution of power density. The transmitters can be single-frequency, multiple-frequency, or broadband. An application of this scenario is for powering multiple sensors whose location is not precisely known.
- Unknown transmitters over a range of frequencies, power levels, generally unpolarized, with varying low-level spatial power densities, such as in energy harvesting or energy scavenging.

The incident power density on the rectenna, $S(\theta, \phi, f, t)$, is a function of incident angles, and can vary over the spectrum and in time. The effective area of the antenna, $A_{\text{eff}}(\theta, \phi, f)$, will be different at different frequencies, for different incident polarizations and incidence angles. The average RF power over a range of frequencies at any instant in time is given by

$$P_{\text{RF}}(t) = \frac{1}{f_{\text{high}} - f_{\text{low}}} \int_{f_{\text{low}}}^{f_{\text{high}}} \int_0^{4\pi} S(\theta, \phi, f, t) A_{\text{eff}}(\theta, \phi, f) d\Omega df$$

The DC power for a single-frequency (f_i) input RF power is given by

$$P_{\text{DC}}(f_i) = P_{\text{RF}}(f_i, t) \cdot \eta(P_{\text{RF}}(f_i, t), \rho, Z_{\text{DC}})$$

where η is the conversion efficiency, and depends on the impedance match $\rho(P_{\text{RF}}, f)$ between the antenna and the rectifier circuit, as well as the DC load impedance. The reflection coefficient in turn is a nonlinear function of power and frequency. Therefore, to find the efficiency of the rectenna, the following measurements can be performed:

- Illuminate the rectenna from a known distance and incidence angle with a known frequency and power; thus $S(\theta, \phi, f, t)$ is known.
- For each frequency of interest, determine the geometrical electrical area of the antenna. This is generally larger than the effective area and therefore overestimates the received RF power.
- Measure the DC power as a function of the DC load impedance (resistance).
- Calculate the estimated conversion efficiency as $P_{\text{RF}}/P_{\text{DC}}$. Since the DC power is measured directly, and the RF power is overestimated, the resulting efficiency will be an underestimate.

This process should be done at each frequency in the range of interest. However, the process is not linear and DC powers obtained in that way cannot be simply added in order to find a multifrequency efficiency.

Narrowband Rectenna Design

For most rectennas and arrays reported to date, the antenna is matched to the diode around one frequency at a well-defined polarization and assuming relatively high incident power levels, e.g. in Brown.⁷⁴ For example, the rectenna shown in Figure 32-24*b* is linearly polarized and designed to operate at 5.8 GHz with an incident power of 50mW corresponding to an incident power density of around 3.2 mW/cm^2 , assuming a dipole effective area of $\lambda^2/8$. In this case, the incident wave carries enough power to turn on the diode, and rectification efficiency can be very high (>80%). A different design is required for low incident power levels. For example, the unlicensed ISM band around 2.4 GHz is an appropriate frequency range for low-power wireless sensor powering. In this scenario, the sensor is mobile and therefore the polarization changes. Thus, the rectenna is designed to be dual-polarized with two rectifier diodes. Each diode rectifies one of the two received polarizations, and the DC signals add. The rectenna is a $19 \times 19\text{-mm}$ square patch, with a $6 \times 6\text{-cm}$ square ground plane on a Rogers Duroid 6010 substrate ($\epsilon_r = 10.2$, thickness = 50 mil) chosen to reduce the antenna size. A Schottky diode is connected at each of the two centers of the two orthogonally polarized patch radiating edges (see Figure 32-25*a*). A via isolated from the patch ground plane terminates each diode to RF/DC ground, and the DC output is taken from the RF short in the center of the patch. Figure 32-25*b* shows a histogram of measured power levels for an incident power level of $170 \mu\text{W/cm}^2$ for different incident polarization combinations, which will all be present in the multipath environment envisioned in this application. This rectenna operates with incident power levels as low as $10 \mu\text{W/cm}^2$ and is capable of powering a low-power wireless sensor.

Broadband Rectenna Design

For low-power applications, as is the case for collected ambient energy, there is generally not enough power to drive the diode in a high-efficiency mode. Furthermore, rectification over multiple octaves requires a different approach from standard matching techniques. In a rectenna application, the antenna itself can be used as the matching mechanism instead of using a transmission-line or lumped-element matching circuit. The antenna design is therefore heavily dependent on the diode characteristics. A source-pull of the diode is performed over a given area of the Smith chart for different input powers, and the resulting DC voltage and

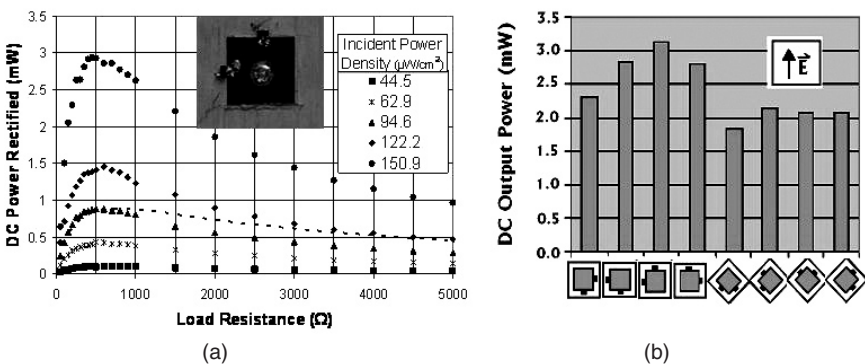


FIGURE 32-25 (a) Measured rectified DC power as a function of the DC load for different incident power levels. A photograph of dual-polarized 2.4-GHz patch rectenna is shown in the inset. (b) Histogram of measured DC power for $170 \mu\text{W/cm}^2$ incident power density and different polarizations of the patch.

is quantified for each source impedance. The region of optimal source impedance is used for optimizing the antenna design. Since the harmonic balance technique calculates the response at all harmonics of the excitation frequency, in the simulation an assumption must be made for the impedance seen by the reflected harmonics, and in the presented case this impedance was set to the impedance of a broadband, self-complementary antenna, $189\ \Omega$. Usually, the optimal source impedance moves counter-clockwise along a constant admittance circle with increasing frequency due to the junction capacitance.

For maximal power transfer, the antenna impedance would match the optimal diode impedance for all frequencies. Since this is difficult to accomplish, a possible suboptimal approach is to present a constant impedance to the diode by using a frequency-independent antenna element. An equiangular spiral was chosen as the array element since it is uniplanar with convenient feedpoint for diode connection, and can be left- or right-hand circularly polarized. A single element was simulated with full-wave CAD tools (Ansoft's Ensemble and Zeland's IE3D) resulting in a one-port frequency-dependent impedance that becomes the diode load in the rectenna. A diode is connected at the antenna feed and the resulting rectenna element performance is shown in Figure 32-26a. The disagreement around 4 GHz is believed to be caused by the 1-cm-long unbalanced coaxial feed, which is not part of the active antenna.

A 64-element array of left- and right-hand circularly polarized spiral elements, each with a rectifier diode, is shown in Figure 32-26b, as described in Hagerty et al.²² The RF powers received independently by each element are summed upon rectification as DC currents and/or voltages. The array was characterized for incident power densities ranging from tens of nW/cm^2 to $0.1\ \text{mW}/\text{cm}^2$. The rectenna array receives radiation from all directions (there is no ground plane), and to measure the re-radiated power at the harmonics, the array was placed halfway between a transmit and receive test antenna, perpendicular to the line-of-sight axis. The DC voltage is measured across a $100\text{-}\Omega$ load. The rectification efficiency reaches the 20% range for an incident power density of $0.1\ \text{mW}/\text{cm}^2$ and arbitrary polarization. The array scaling in size is trivial, since there is no RF feed network, and only the DC output network is required. If space allows, this type of array can be made in the form of wallpaper for collecting as much power as possible. The DC collecting lines can be made to be reconfigurable, so that either current or voltage is summed, depending on the DC load.

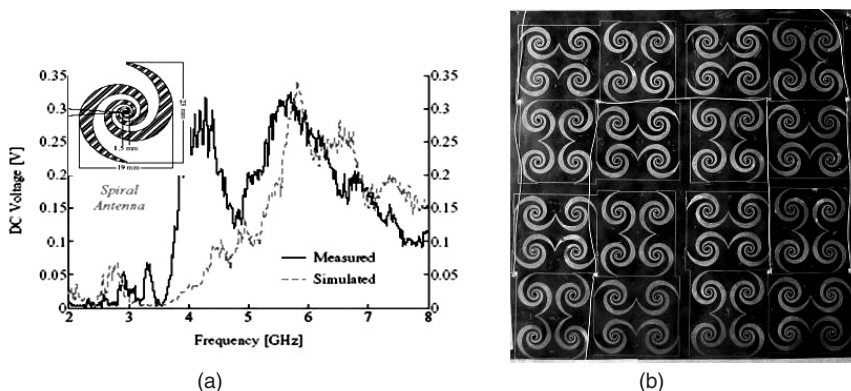


FIGURE 32-26 (a) Nonlinear harmonic-balance simulation and measurements of the DC rectified voltage response across $60\ \Omega$ for the spiral antenna in the inset, over a 4:1 bandwidth (after J. A. Hagerty et al.²² © IEEE 2004); (b) photograph of 64-element dual-circularly polarized 2- to 18-GHz spiral rectenna array

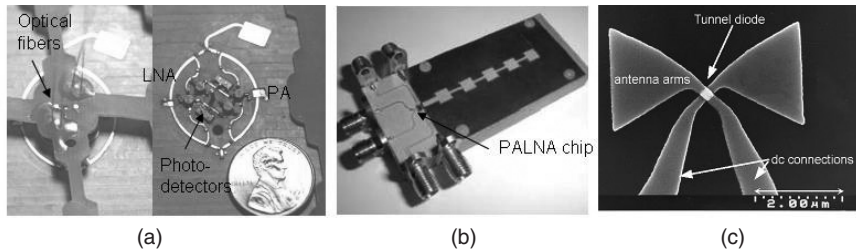


FIGURE 32-27 (a) Unit cell of a 36-element T/R optically controlled active array. The optical power routes the signal between a receive-mode LNA and a transmit-mode PA. (b) Photograph of a switchless simplex T/R series-fed amplifier patch array at 24 GHz. (c) A 4- μm -long infrared imaging antenna and detector.

32.7 OTHER ACTIVE ANTENNAS

The goal of this section is to briefly overview some interesting active antennas that do not fit any of the categories discussed in Sections 2 to 6, but have found useful applications or have promise for future front ends. Other design methodologies⁷⁵ for antenna-circuit modules are not discussed here. Figure 32-27a shows a unit cell of a 10-GHz optically switched transmit-receive (T/R) multibeam antenna array.⁷⁶ Two pairs of multimode optical fibers are held by an FR4 mount and the light from a laser diode is coupled to off-the-shelf photodetectors, which control the bias of pin-diode SPDT switches. The signal between two patch antennas in each unit cell of the amplifier array is routed either through an LNA or PA at nsec speed with only pJ of optical energy, and the optical fibers are transparent at 10 GHz and do not affect the radiation pattern. Figure 32-27b shows a single T/R series-fed 24-GHz patch array fed by a bi-directional (PALNA) chip amplifier, where the T/R path is controlled by bias, eliminating switch loss.⁷⁷ Figure 32-27c shows an infrared antenna with a tunnel-diode detector, and other similar extremely high-frequency antennas have been demonstrated for millimeter-wave and THz imaging applications.^{78,79}

REFERENCES

1. R. J. Mailloux, *Phased Array Antenna Handbook*, 2nd Ed. (Norwood, MA: Artech House Publishers, 2005).
2. R. C. Hansen, *Phased Array Antennas* (New York: John Wiley & Sons, Inc., 1998).
3. Q. Dongjiang, R. Molino, S. M. Lardizabal, B. Pillans, P. M. Asbeck, and G. Jerinic, "An Intelligently Controlled RF Power Amplifier with a Reconfigurable MEMS-varactor Tuner," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 3, part 2 (March 2005): 1089–1095.
4. M. Forman and Z. Popovic, "A Tunable Second-Resonance Cross-Slot Antenna," *1997 IEEE AP-S Int. Symp. Digest* (1997): 18–21.
5. N. Behdad and K. Sarabandi, "A Varactor-Tuned Dual-Band Slot Antenna," *IEEE Trans. Antennas and Propagation*, vol. 54, no. 2 (February 2006): 401–408.
6. K. Chang, K. Hummer, and G. Gopalakrishnan, "Active Radiating Element Using FET Source Integrated with Microstrip Patch Antenna," *Electronics Letters*, vol. 24, no. 21 (October 1988): 1347–1348.
7. R. H. Rasshofer and E. M. Biebl, "A Direction Sensitive, Integrated, Low Cost Doppler Radar Sensor for Automotive Applications," *1998 IEEE MTT-S Int. Microwave Symp. Digest* (1998): 1055–1058.

8. Z. Popovic, R. Weikle, M. Kim, and D. Rutledge, "A 100-MESFET Planar Grid Oscillator," *IEEE Trans. Microwave Theory Tech.*, vol. 39, no. 2 (February 1991): 193–200.
9. J. Birkeland and T. Itoh, "A 16 Element Quasi-Optical FET Oscillator Power Combining Array with External Injection Locking," *IEEE Trans. Microwave Theory Tech.*, vol. 40, no. 3 (March 1992): 475–481.
10. A. Mortazawi and B. DeLoach, "Spatial Power Combining Oscillators Based on an Extended Resonance Technique," *IEEE Trans. Microwave Theory Tech.*, vol. 42, no. 12 (December 1994): 2222–2228.
11. P. Liao and R. A. York, "A New Phase-Shifterless Beam-Scanning Technique Using Arrays of Coupled Oscillators," *IEEE Trans. Microwave Theory Tech.*, vol. 41, no. 10 (October 1993): 1810–1815.
12. S. Hollung, A. E. Cox, and Z. Popovic, "A Bi-directional Quasi-Optical Lens Amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. 45, no. 12, part 2 (December 1997): 2352–2357.
13. S. Nam, T. Uwano, and T. Itoh, "Microstrip-Fed Planar Frequency-Multiplying Space Combiner," *IEEE Trans. Microwave Theory Tech.*, vol. 35, no. 12 (December 1987).
14. K. Chang, R. A. York, P. S. Hall, and T. Itoh, "Active Integrated Antennas," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 3 (March 2002): 937–944.
15. J. B. Hacker, R. M. Weikle, M. Kim, M. P. De Lisio, and D. B. Rutledge, "A 100-Element Planar Schottky Diode Grid Mixer," *IEEE Trans. Microwave Theory Tech.*, vol. 40, no. 3 (March 1992): 557–562.
16. A. Luukanen, E. N. Grossman, A. J. Miller, P. Helisto, J. S. Penttila, H. Sipola, and H. Seppa, "An Ultra-Low Noise Superconducting Antenna-Coupled Microbolometer with a Room-Temperature Read-Out," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 8 (August 2006): 464–466.
17. C. W. Pobanz and T. Itoh, "A Conformal Retrodirective Array for Radar Applications Using a Heterodyne Phased Scattering Element," 1995 *IEEE MTT-S Int. Microwave Symp. Digest*, vol. 2, (1995): 905–908.
18. R. J. Hwu, C. F. Jou, W. W. Lam, U. Lieneweg, D. C. Steit, N. C. Luhmann, J. Maserjian, and D. B. Rutledge, "Watt-Level Millimeterwave Monolithic Diode-grid Frequency Multipliers," 1988 *IEEE MTT-S Int. Microwave Symp. Digest*, vol. 1 (1988): 533–536.
19. B. Strassner and K. Chang, "Passive 5.8-GHz Radio-Frequency Identification Tag for Monitoring Oil Drill Pipe," *IEEE Trans. Microwave Theory Tech.*, vol. 51, no. 2, part 1 (February 2003): 356–363.
20. C. Walsh, S. Rondineau, M. Jankovic, G. Zhao, and Z. Popovic, "A Conformal 10-GHz Rectenna for Wireless Powering of Piezoelectric Sensor Electronics," *IEEE MTT IMS Digest* (June 2005): 143–146.
21. P. Koert and J. T. Cha, "Millimeter Wave Technology for Space Power Beaming," *IEEE Trans. Microwave Theory Tech.*, vol. 40, no. 6 (June 1992): 1251–1258.
22. J. A. Hagerty, F. B. Helmbrecht, W. H. McCalpin, R. Zane, and Z. Popovic, "Recycling Ambient Microwave Energy with Broad-band Rectenna Arrays," *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 3 (March 2004): 1014–1024.
23. G. D. Boreman, C. Fumeaux, W. Herrmann, F. K. Kneubuhl, and H. Rothuizen, "Tunable Polarization Response of a Planar Asymmetric Spiral Infrared Antenna," *Optics Letters*, vol. 23, no. 24 (December 1998): 1912–1914.
24. J. Vian and Z. Popovic, "A Transmit/Receive Active Antenna with Fast Low-Power Optical Switching," *IEEE Trans. Microwave Theory Tech.*, vol. 48, no. 12 (December 2000): 2686–2691.
25. K. D. Stephan and T. Itoh, "A Planar Quasi-Optical Subharmonically Pumped Mixer Characterized by Isotropic Conversion Loss," *IEEE Trans. Microwave Theory Tech.*, vol. 32, no. 1 (January 1984): 97–102.
26. M. Weiss and Z. Popovic, "A 10 GHz High-Efficiency Active Antenna," 1999 *IEEE MTT-S Int. Microwave Symp. Digest*, vol. 2 (1999): 663–666.

27. Z. B. Popovic and R. A. York, *Active and Quasi-Optical Arrays for Solid-State Power Combining* (New York: John Wiley & Sons, Inc., 1997).
28. R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook* (Norwood, MA: Artech House Publishers, 2001).
29. S. H. Al-Charchafchi and M. Frances, "Electronically Tunable Microstrip Patch Antennas," *1998 IEEE AP-S Int. Symp. Digest*, vol. 1 (June 1998): 304–307.
30. J. Ollikainen, O. Kivekas, and P. Vainikainen, "Low-loss Tuning Circuits for Frequency-Tunable Small Resonant Antennas Personal, Indoor and Mobile Radio Communications, 2002," *The 13th IEEE International Symposium*, vol. 4 (September 2002): 1882–1887.
31. J. A. Navarro, Y. Shu, and K. Chang, "Wideband Integrated Varactor-Tunable Active Notch Antennas and Power Combiners," *1991 IEEE MTT-S Int. Microw. Symp. Digest*, vol. 3 (1993): 1257–1260.
32. D. Peroulis, K. Sarabandi, and L. Katehi, "Design of Reconfigurable Slot Antennas," *IEEE Trans. Antennas and Propagation*, vol. 53, no. 2 (February 2005): 645–654.
33. J. Roscoe, L. Shafai, A. Ittipiboon, M. Cuhaci, and R. Douville, "Tunable Dipole Antennas," *1993 IEEE AP-S Int. Symp. Digest* (1993): 672–675.
34. D. Anagnostou, G. Zheng, M. Cryssomallis, J. Lyke, G. Ponchak, J. Papapolymerou, and C. Christodoulou, "Design, Fabrication and Measurements of an RF-MEMS-Based Self-Similar Reconfigurable Antenna," *IEEE Trans. Antennas and Prop.*, vol. 54, no. 2 (February 2006): 422–433.
35. A. M. Castro-Vilaro and R. A. Rodriguez-Solis, "Tunable Folded-Slot Antenna with Thin Film Ferroelectric Material," *2003 IEEE AP-S Int. Symp. Digest*, vol. 2 (June 2003): 549–552.
36. A. Mortazawi, T. Itoh, and J. Harvey, *Active Antennas and Quasi-Optical Arrays* (New York: IEEE, Inc., 1999).
37. B. K. Kormanyos, S. E. Rosenbaum, L. P. Katehi, and G. M. Rebeiz, "Monolithic 155 GHz and 215 GHz Quasi-Optical Slot Oscillators," *1994 IEEE MTT-S Int. Microwave Symp. Digest* (1994): 835–838.
38. J. A. Hagerty and Z. Popovic, "A 10 GHz Active Annular Ring Antenna," *2002 IEEE AP-S Int. Symp. Digest*, vol. 2 (June 2002): 284–287.
39. T. Perkins, "Microstrip Patch Antenna with Embedded Impatt Diode Oscillator," *1986 IEEE AP-S Int. Symp. Digest*, vol. 24 (June 1986): 447–450.
40. M. J. Wengler, B. Guan, and E. K. Track, "190-GHz Radiation from a Quasioptical Josephson Junction Array," *IEEE Trans. Microwave Theory Tech.*, vol. 43, no. 4, part 1 (April 1995): 984–988.
41. M. M. Kaleja, A. J. Herb, R. H. Rasshofer, G. Friedsam, and E. M. Biebl, "Imaging RFID System at 24 GHz for Object Localization," *1999 IEEE MTT-S Int. Microwave Symp. Digest*, vol. 4 (June 1999): 1497–1500.
42. M. M. Kaleja, "Active Integrated Antennas for Sensor and Communication Applications," Ph.D. dissertation, Technische Universität München, 2001.
43. J. B. Hacker, M. P. De Lisio, K. Moonil, L. Cheh-Ming, L. Shi-Jie, S. W. Wedge, and D. B. Rutledge, "A 10-W X-band Grid Oscillator," *1994 IEEE MTT-S Int. Microwave Symp. Digest*, vol. 2 (1994): 823–826.
44. J. Bae, T. Unou, T. Fujii, and M. Mizuno, "Spatial Power Combining of Gunn Diodes Using an Overmoded-Waveguide Resonator at Millimeter Wavelengths," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 12 (December 1998): 2289–2294.
45. R. A. York and T. Itoh, "Injection- and Phase-Locking Techniques for Beam Control," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 11 (November 1998): 1920–1929.
46. W. Duerr, W. Menzel, and H. Schumacher, "A Low-Noise Active Receiving Antenna Using a SiGe HBT," *IEEE Microwave and Guided Wave Lett.*, vol. 7, no. 3 (March 1997): 63–67.
47. D. Radulovic, A. Nesic, and I. Radnovic, "Impedance of Patch Antenna for Active Antenna Structures," *IEEE AP International Symposium Digest* (July 2004): 3931–3934.
48. J. Luine L. Abelson, D. Brundrett, J. Burch, E. Dantsker, K. Hummer, G. Kerber, M. Wire, K. Yokoyama, D. Bowling, M. Nee1, S. Hubbell, and K. Li, "Application of a DC SQUID Array

- Amplifier to an Electrically Small Active Antenna," *IEEE Trans. On Applied Superconductivity*, vol. 9, no. 2 (June 1999): 4141–4144.
49. H. An, B. Nauwelaers, A. Van de Capelle, and R. Bosio, "A Novel Measurement Technique for Amplifier-type Active Antennas," *IEEE MTT-S Int. Microwave Symp. Dig.* (June 1994): 1473–1476.
 50. J. Peeters Weem and Z. Popovic, "A Method for Determining Noise Coupling in a Phased Array Antenna," *2001 IEEE MTT IMS Digest* (May 2001): 271–274.
 51. S. Pajic and Z. Popovic, "An Efficient 16-element X-band Spatial Combiner of Switched-mode Power Amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 51, no. 73 (July 2003): 1863–1870.
 52. J. Hubert, J. Schoenberg, and Z. Popovic "A Ka-band Quasi-optical Amplifier," *1995 IEEE MTT-S Int. Symp. Dig.* (May 1995): 585–588.
 53. T. Marshall, M. Forman, and Z. Popovic, "Two Ka-band Quasi-Optical Amplifier Arrays," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 12 (December 1999): 2568–2573.
 54. N-S. Cheng, A. Alexanian, M. Case, D. Rensch, and R. A. York, "40-W CW Broadband Spatial Power Combiner Using Dense Finline Arrays," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 7 (July 1999): 1070–1076.
 55. N-S. Cheng, P. Jia, D. Rensch, and R. A. York, "A 120-W X-Band Spatially Power Combined Solid-State Amplifier," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 12 (December 1999): 2557–2562.
 56. R. M. Weikle, M. Kim, J. B. Hacker, M. P. DeLisio, Z. Popovic, and D. Rutledge, "Transistor Oscillator and Amplifier Grids," *Invited paper, Proc. IEEE*, vol. 80, no. 11 (November 1992): 1800–1809.
 57. B. Deckman, D. Deakin, E. Sovero, and D. Rutledge, "A 5-W, 37-GHz Monolithic Grid Amplifier," *IEEE MTT-S Int. Microwave Symp. Dig.* (2000): 805–808.
 58. M. P. DeLisio, B. Deckman, C.-T. Cheung, S. C. Martin, D. P. Nakhla, E. E. Hartmann, C. J. Rollison, J. B. Pacetti, and J. Rosenberg, "A Ka-band Grid Amplifier Module with over 10 Watts Output Power," *IEEE MTT-S Int. Microwave Symp. Dig.*, (2004): 83–86.
 59. C.-T. Cheung, R. Tsai, R. Kagiwada, and D. B. Rutledge, "V-band Transmission and Reflection Grid Amplifier Packaged in Waveguide," *IEEE MTT-S Int. Microwave Symp. Dig.* (2003): 1863–1866.
 60. S.C. Ortiz, J. Hubert, L. Mirth, E. Schlecht, and A. Mortazawi, "A High-Power Ka-Band Quasi-Optical Amplifier Array," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 2 (February 2002): 487–494.
 61. C. F. Jou, W. W. Lam, H. Z. Chen, K. S. Stolt, N. C. Luhmann, Jr., and D. B. Rutledge, "Millimeter-wave Diode Grid Frequency Doubler," *IEEE Trans. Microwave Theory Tech.*, vol. 36, no. 11 (Nov. 1988): 1507–1514.
 62. A. Moussessian, M. Wanke, Y. Li, J-C. Chiao, S. J. Allen, T. Crowe, and D. B. Rutledge, "A Terahertz Grid Frequency Doubler," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 11 (November 1998): 1976–1981.
 63. M. Sironen, Y. Qian, and T. Itoh, "A Subharmonic Self-Oscillating Mixer with Integrated Antenna for 60-GHz Wireless Applications," *IEEE Trans. Microwave Theory Tech.*, vol. 49, no. 3 (March 2001): 442–450.
 64. J. B. Hacker, R. M. Weikle, M. Kim, M. P. De Lisio, and D. B. Rutledge, "A 100-Element Planar Schottky Diode Grid Mixer," *IEEE Trans. Microwave Theory Tech.*, vol. 40, no. 3 (March 1992): 557–562.
 65. C. Pobanz and T. Itoh, "A Conformal Retrodirective Array for Heterodyne Applications Using a Heterodyne Phase Scattering Element," *IEEE MTT-S International Microwave Symposium Digest*, (1995): 905–908.
 66. Jim Glenn (ed.), *The Complete Patents of Nikola Tesla* (New York: Barnes and Noble Books, 1994): 346–360.
 67. W. C. Brown, "The History of Power Transmission by Radio Waves," *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 9 (Sept. 1984): 1230–1242.

68. N. Shinohara and H. Matsumoto, "Experimental Study of Large Rectenna Array for Microwave Energy Transmission," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, no. 3 (March 1998): 261–267.
69. T. Yoo and K. Chang, "Theoretical and Experimental Development of 10 and 35 GHz Rectennas," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 6 (June 1992): 1259–1266.
70. L. W. Epp, A. R. Khan, H. K. Smith, and R. P. Smith, "A Compact Dual-Polarized 8.51-GHz Rectenna for High-Voltage (50 V) Actuator Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 1 (January 2000): 111–120.
71. Y. Fujino, T. Ito, M. Fujita, N. Kaya, H. Matsumoto, K. Kawabata, H. Sawada, and T. Onodera, "A Driving Test of a Small DC Motor with a Rectenna Array," *IEICE Trans. Commun.*, vol. E77-B, no. 4 (April 1994): 526–528.
72. J. O. McSpadden, R. M. Dickinson, L. Fan, and Kai Chang, "Design and Experiments of a High-Conversion-Efficiency 5.8-GHz Rectenna," *IEEE MTT IMS Digest*, vol. 2 (1998): 1161–1164.
73. B. Strassner and K. Chang, "A Circularly Polarized Rectifying Antenna Array for Wireless Microwave Power Transmission with over 78% Efficiency," *IEEE MTT IMS Digest* (2002): 1535–1538.
74. W. C. Brown, "An Experimental Low Power Density Rectenna," *IEEE MTT-S International Microwave Symposium Digest* (1991): 197–200.
75. K. C. Gupta and P. Hall (eds.), *Analysis and Design of Integrated Circuit Antenna Modules* (New York: John Wiley and Sons, Inc., 2000).
76. J. Vian and Z. Popovic, "A Transmit/Receive Active Antenna with Fast Low-power Optical Switching," *IEEE Trans. Microwave Theory Tech.*, vol. 48, no. 12 (December 2000): 2686–2691.
77. D. Lu, D. Rutledge, M. Kovacevic, and J. Hacker, "A 24-GHz Patch Array with a Power Amplifier/Low-Noise Amplifier MMIC," *Intl. Journal of Infrared and Millimeter Waves*, vol. 23 (May 2002): 693–704.
78. G. Boreman, C. Fumeaux, W. Herrman, F. Kneubuhl, and H. Rothuizen, "Tunable Polarization Response of a Planar Asymmetric-spiral Infrared Antenna," *Optics Letters*, vol. 23, no. 24 (December 1998): 1912–1914.
79. C. Dietlein, J. D. Chisum, M. D. Ramírez, E. N. Grossman, A. Luukanen, and Z. Popovic, "Integrated Microbolometer Antenna Characterization from 109–650 GHz," *2007 IEEE MTT IMS Digest* (2007).

Chapter 33

Fractal Antennas

Douglas H. Werner
Joshua S. Petko
Thomas G. Spence
The Pennsylvania State University

CONTENTS

33.1	INTRODUCTION	33-2
33.2	USEFUL FRACTAL ANTENNA GEOMETRIES	33-2
33.3	ITERATED FUNCTION SYSTEMS	33-6
33.4	FRACTAL ANTENNA ELEMENTS	33-9
33.5	FRACTAL ANTENNA ARRAYS	33-13
33.6	ANTENNA ARRAYS BASED ON FRACTAL AND APERIODIC TILINGS	33-21

33.1 INTRODUCTION

The term *nature-based antenna design* was recently coined in Neiss et al¹ to describe the process of designing complex antenna systems based on concepts inspired by nature. Examples of nature-based techniques that have been recently applied to solve advanced antenna engineering design problems include fractal geometry, neural networks, fuzzy logic, genetic algorithms, particle swarm optimization, and ant colony optimization. In some cases, even two or more of these nature-based design techniques have been combined together to effectively solve a specific antenna design problem. An example of this can be found in Petko and Werner,² where a powerful nature-based antenna array synthesis technique was introduced that incorporates aspects of both fractal geometry and genetic algorithms.

Traditional approaches to the analysis and design of antenna systems have their foundation in Euclidean geometry. In this chapter, however, we focus on nature-based antenna design concepts that employ fractal geometry. The term *fractal*, meaning broken or irregular fragments, was originally used by Mandelbrot³ to describe a family of complex shapes that possess an inherent self-similarity or self-affinity in their geometrical structure. The original inspiration for the development of fractal geometry came largely from an in-depth study of the patterns of nature. Fractals are abundant in nature, with a few examples of natural fractals being snowflakes, ferns, trees, coastlines, mountain ranges and even galaxies.⁴

Several book chapters and review articles have been published that provide a comprehensive overview of research in the area of fractal antenna engineering.⁵⁻⁹ This research can be divided into two categories, the first of which deals with the study of fractal-shaped antenna elements, and the second of which concerns the use of fractals in the design of antenna arrays. The next section presents a brief summary of some of the most popular types of fractal geometries found to be useful in antenna engineering applications. Section 33.3 then provides an introduction to the theory of iterated function systems (IFSs) along with a discussion of the important role they play in the mathematical description of fractals. Sections 33.4 and 33.5 summarize approaches that employ fractal concepts in the design of antenna elements and arrays, respectively. Finally, Section 33.6 is devoted to design methodologies for antenna arrays that are based on fractal and aperiodic tilings.

33.2 USEFUL FRACTAL ANTENNA GEOMETRIES

This section presents a brief introduction to some of the more common fractal geometries that have been found to be useful in antenna engineering applications. This includes applications involving fractal-shaped antenna elements as well as conventional Euclidean antenna elements placed in fractal array configurations. The first and perhaps the most recognizable fractal geometry that will be considered is known as the Sierpinski gasket.⁴ Figure 33-1 illustrates the first few stages in the construction of the Sierpinski gasket fractal. The procedure for geometrically constructing this fractal begins with an equilateral triangle contained in the plane, shown in black as stage 1 of Figure 33-1. The next step in the construction process is to remove the central triangle whose vertices are located at the midpoints of the sides of the original stage 1 triangle, which leads to the geometry shown in stage 2 of Figure 33-1. This process is then repeated for the three remaining black triangles, with the result shown in stage 3 of Figure 33-1. The next stage (stage 4) in the construction of the Sierpinski gasket is also shown in Figure 33-1. The Sierpinski gasket fractal is

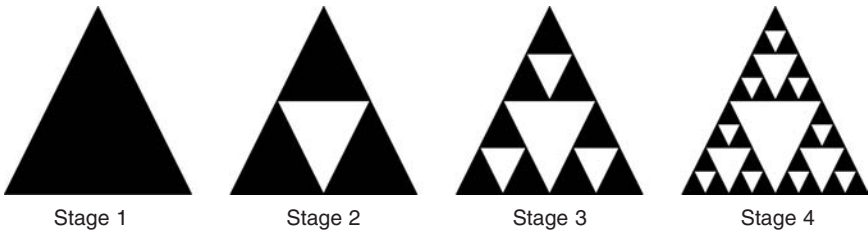


FIGURE 33-1 The first four stages in the generation of a Sierpinski gasket fractal

generated by carrying out this iterative process an infinite number of times. It follows from this definition that the Sierpinski gasket is an example of a self-similar fractal; i.e., it is composed from small but exact copies of itself. This self-similar property of the Sierpinski gasket fractal is illustrated in Figure 33-2.

Another useful fractal is commonly referred to as the Koch snowflake.⁴ This fractal also starts out as an equilateral triangle in the plane, as depicted by the solid black triangle in stage 1 of Figure 33-3. However, unlike the Sierpinski gasket, which was formed by systematically removing smaller and smaller triangles from the original structure, the Koch snowflake is constructed by adding smaller and smaller triangles to the structure in an iterative fashion. The first four stages in the iterative process of constructing the Koch snowflake are illustrated in Figure 33-3.

A number of structures based on purely deterministic or random fractal trees have also proven to be extremely useful in developing new design methodologies for antenna elements and array configurations. An example of a deterministic fractal tree is shown in Figure 33-4a. A ternary (three-branch) generator is used for the first three stages of growth. A fractal-random tree, on the other hand, is a type of fractal tree which uses multiple generators selected in random order to form the tree structure. Figure 33-4b shows the first three stages of growth of a random fractal tree. In this case the tree is grown by randomly selecting between two generators, one with three branches and the other with two branches. Fractal-random trees are more representative of natural trees or plants than deterministic fractal trees, which appear too ordered. The examples presented in Figure 33-4 are of two-dimensional fractal trees. However, the concept of a fractal tree can be easily generalized to three-dimensions. Figure 33-5 shows the first four stages in the growth of a three-dimensional deterministic fractal tree.^{8,10}

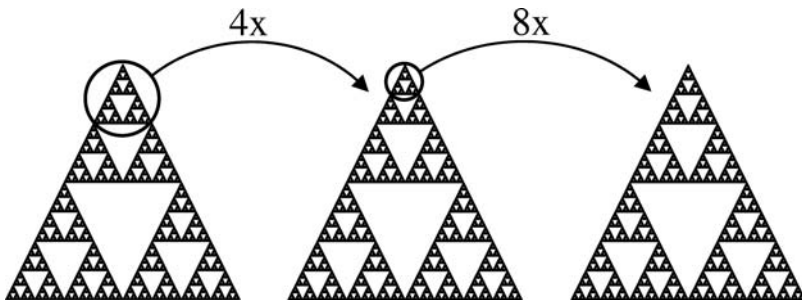


FIGURE 33-2 Illustration of the self-similar geometry of the Sierpinski gasket fractal

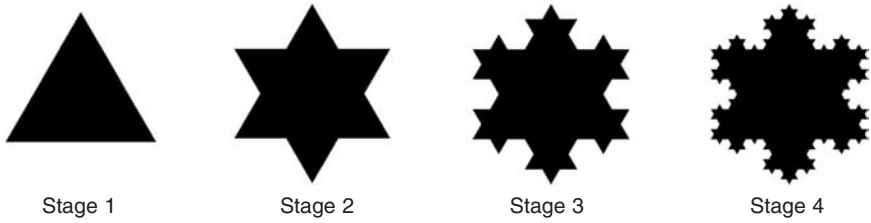


FIGURE 33-3 The first four stages in the generation of a Koch fractal snowflake

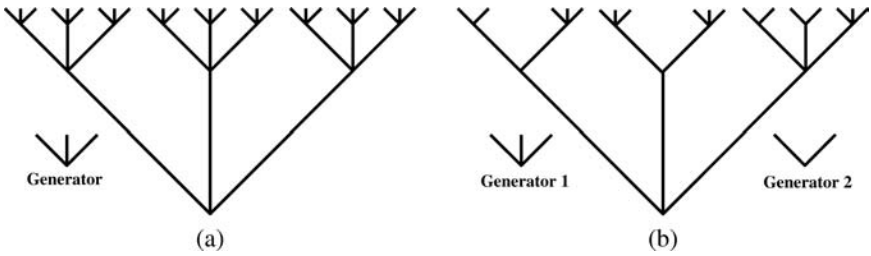


FIGURE 33-4 Example of (a) deterministic fractal tree and (b) fractal-random tree

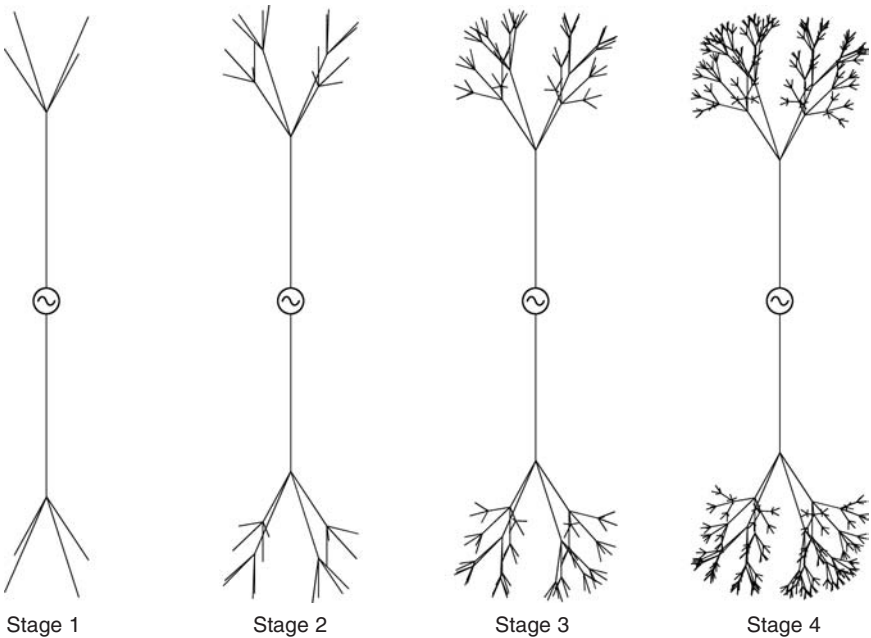


FIGURE 33-5 First four stages of growth of a three-dimensional fractal tree with a four-branch generator (after Petko and Werner¹⁰ © IEEE 2004)

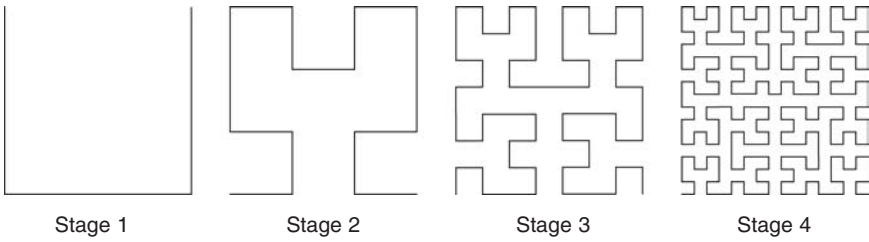


FIGURE 33-6 The first four stages of growth for a space-filling Hilbert curve

The space-filling properties of the Hilbert curve and related curves make them attractive candidates for use in the design of fractal antennas. The first four stages in the construction of the Hilbert curve are shown in Figure 33-6.^{4,11} The Hilbert curve is an example of a space-filling fractal curve that is self-avoiding (i.e., has no intersection points). Another important example of a space-filling curve with applications in antenna engineering is the Peano-Gosper curve.^{3,11} Figure 33-7 illustrates the first three stages in the construction of the Peano-Gosper curve. The initiator is shown in Figure 33-7a as the dashed-line segment superimposed on the stage 1 generator. A copy of the stage 1 generator is shown again in Figure 33-7b as the dashed curve. Also shown in Figure 33-7b is the stage 2 Peano-Gosper curve, which is obtained by replacing each of the seven segments in the stage 1 generator by an appropriately scaled copy of itself. The next iteration (stage 3) of the Peano-Gosper curve is shown in Figure 33-7c. At this stage, the space-filling property of the curve starts to become more apparent.

Fractal tiles, or fractiles, represent another important category of fractal objects that has found application in antenna theory and design.^{3,12} Fractiles are tiles or islands that possess fractal boundaries. They are a unique subset of all possible tile geometries that can be used to cover the plane without overlapping or leaving any gaps in between the tiles. We consider here two specific types of fractiles, known as fudgeflakes and Gosper islands. Both fudgeflakes and Gosper islands can be used to cover the plane via a tiling.^{3,12} The first through sixth iterations of the fudgeflake are shown in Figure 33-8, while the first through

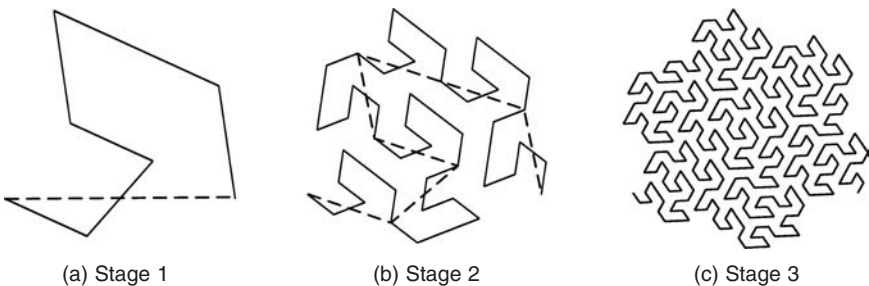


FIGURE 33-7 The first three stages of the space-filling Peano-Gosper curve. The initiator is shown as the dashed-line segment superimposed on the stage 1 generator, which is shown as the dashed curve superimposed on the stage 2 generator.

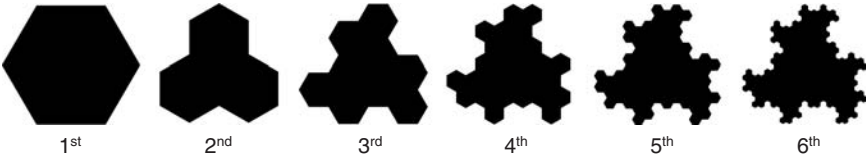


FIGURE 33-8 Geometry of the first through sixth iterations of the fudgeflake fractile

fifth iterations of the Gosper island are shown in Figure 33-9. Finally, Figure 33-10 demonstrates how fudgeflake and Gosper island fractiles can be used to provide a perfect covering of the plane (i.e. no gaps or overlaps between tiles).

33.3 ITERATED FUNCTION SYSTEMS

Iterated function systems (IFSs) are the mathematical language of fractals. They provide a unified approach to the theory of fractal geometry and represent an extremely versatile tool for conveniently generating a wide variety of useful fractal structures.^{4,13} These IFSs are based on a collection of contractions achieved through the application of a series of affine transformations w defined as

$$w \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \quad (33-1)$$

or equivalently as

$$w(x, y) = (ax + by + e, cx + dy + f) \quad (33-2)$$

where a, b, c, d, e , and f are real numbers. Hence, the affine transformation w is represented by six parameters, which may be expressed using the compact notation

$$\begin{pmatrix} a & b & e \\ c & d & f \end{pmatrix} \quad (33-3)$$

such that a, b, c , and d control rotation and scaling, while e and f govern linear translation.

Now if we let w_1, w_2, \dots, w_N be a set of affine linear transformations and A be the initial geometry, then a new geometry can be formed by applying the set of transformations to

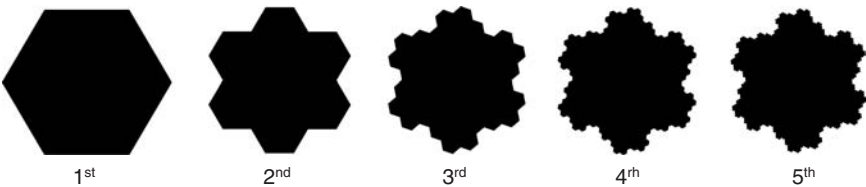


FIGURE 33-9 Geometry of the first through fifth iterations of the Gosper island fractile

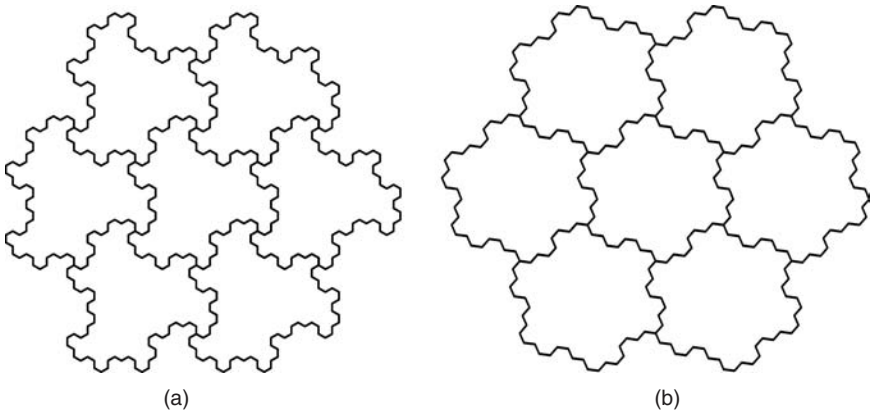


FIGURE 33-10 A covering of the plane by (a) fudgeflake fractiles and (b) Gosper island fractiles

the original geometry A and collecting together the results from $w_1(A)$, $w_2(A)$, ..., $w_N(A)$ in the following way:

$$W(A) = \bigcup_{n=1}^N w_n(A) \quad (33-4)$$

where W is known as the Hutchinson operator.⁴ A fractal geometry can be obtained by repeatedly applying W to the previous geometry in an iterative fashion. For example, if the set A_0 represents the initial geometry, then this iterative process would yield a sequence of Hutchinson operators given by

$$A_1 = W(A_0), A_2 = W(A_1), \dots, A_{k+1} = W(A_k) \quad (33-5)$$

An IFS generates a sequence that converges to a final image A_∞ in such a way that

$$W(A_\infty) = A_\infty \quad (33-6)$$

This image is called the attractor of the IFS and represents a “fixed point” of W , where the “points” in this case are actually defined as sets.

The IFS procedure for generating the well-known Koch fractal curve is demonstrated in Figure 33-11. In this case, the initial set A_0 is the line interval of unit length such that $A_0 = \{x: x \in [0,1]\}$. Four affine linear transformations (i.e., $N=4$) are then applied to A_0 as depicted in Figure 33-11. Next, the results of these four linear transformations are combined together using Eq. 33-4 to form the first iteration of the Koch curve denoted by A_1 . The second iteration of the Koch curve A_2 may then be obtained by applying the same four affine transformations to A_1 and again using Eq. 33-4 to combine the results. Higher-order versions of the Koch curve are generated by simply repeating the iterative process until the desired resolution is achieved. The first four iterations of the Koch curve are shown in Figure 33-12. This sequence of curves eventually converges to the actual Koch fractal (represented by A_∞) as the number of iterations approaches infinity.

IFSs have proven to be a very powerful tool in the analysis and design of fractal antennas. This is primarily because they provide a general framework for the description, classification, and manipulation of fractals.¹³ To further illustrate this important point, an IFS approach was used to produce the Sierpinski gasket shown in Figure 33-13a and the tree-like structure shown in Figure 33-13b, two very different fractal objects. The IFS codes for generating these two objects have also been provided in the figure.⁴

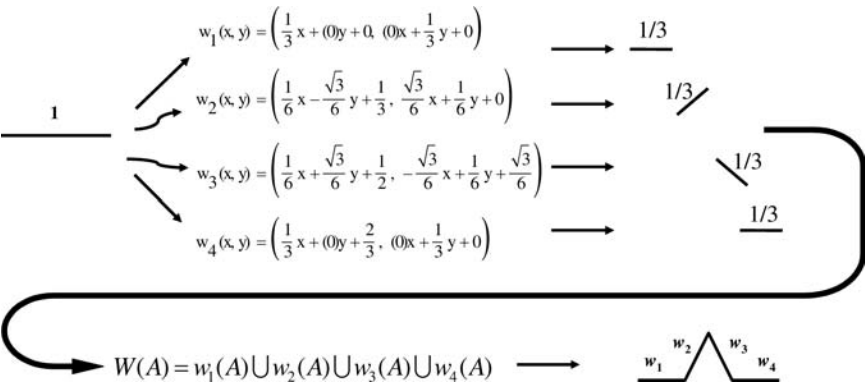


FIGURE 33-11 The standard Koch fractal curve as an IFS (after Werner and ganguly⁹ © IEEE 2003)

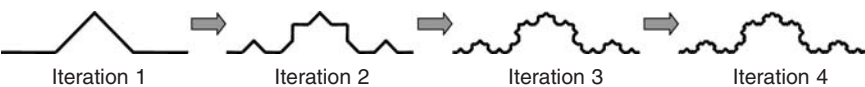


FIGURE 33-12 The first four stages in the generation of the standard Koch fractal curve via an IFS approach

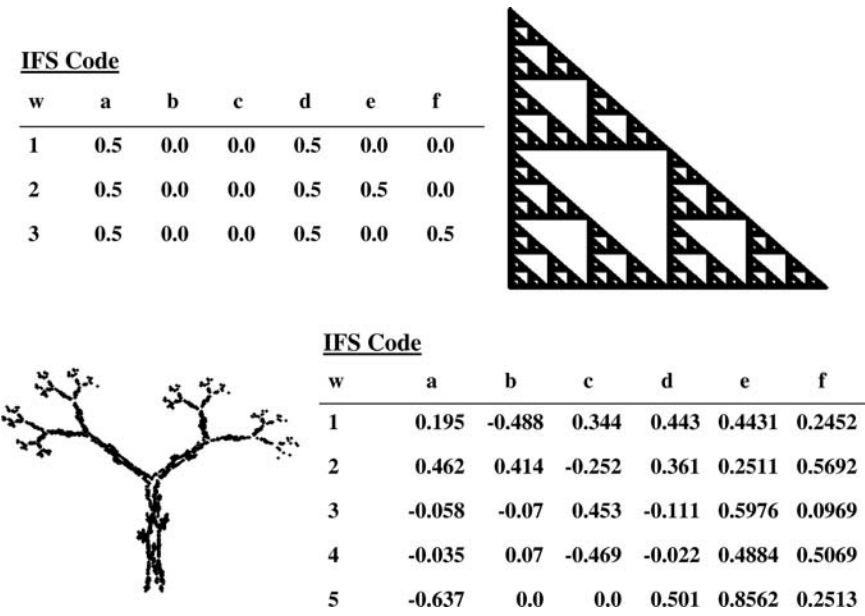


FIGURE 33-13 The IFS code for a Sierpinski gasket and a fractal tree (after Werner and Ganguly⁹ © IEEE 2003)

33.4 FRACTAL ANTENNA ELEMENTS

Some of the more common fractal geometries that have found applications in antenna engineering are depicted in Figure 33-14. The Koch snowflakes and islands have been primarily used to develop new designs for miniaturized loop as well as microstrip patch antennas.^{8,9,14,15} Moreover, fractile concepts have also been employed in the design of fudgeflake and Gosper island microstrip antennas.¹⁶ The self-similar structure of Sierpinski gaskets and carpets has been exploited to develop multiband and broadband antenna elements.^{7,9,17,18,19,20} Finally, novel designs for miniaturized dipole and monopole antennas have been developed based on a variety of Koch curves and fractal trees.^{7,8,9,10,21,22,23}

The self-similar properties of fractal trees have been utilized to create antenna elements with multiband characteristics.²⁴⁻²⁷ In addition, fractal tree geometries have been used as an effective means to miniaturize antenna designs. A set of dipole antennas has been studied that utilizes three-dimensional fractal tree structures as end-loads in Gianvittorio and Rahmat-Samii.⁸ This set of antennas, illustrated in Figure 33-5, has been shown to exhibit lower resonant frequencies than standard dipole antennas of comparable length while not significantly affecting the bandwidth or the gain. An investigation of performance trends of various three-dimensional end-loaded fractal tree dipoles has also been carried out in Petko and Werner.¹⁰ It has been shown that denser fractal tree end-loads exhibit greater reduction in resonant frequency than less dense structures. The influence of branch elevation angles of the fractal tree generator was also studied in Petko and Werner.¹⁰ Figure 33-15 plots the value of the VSWR that corresponds to the resonant frequency of various four-branch fractal tree antenna designs with different branch elevation angles. The results show that fractal tree antennas with small branch elevation angles have a lower VSWR but a higher

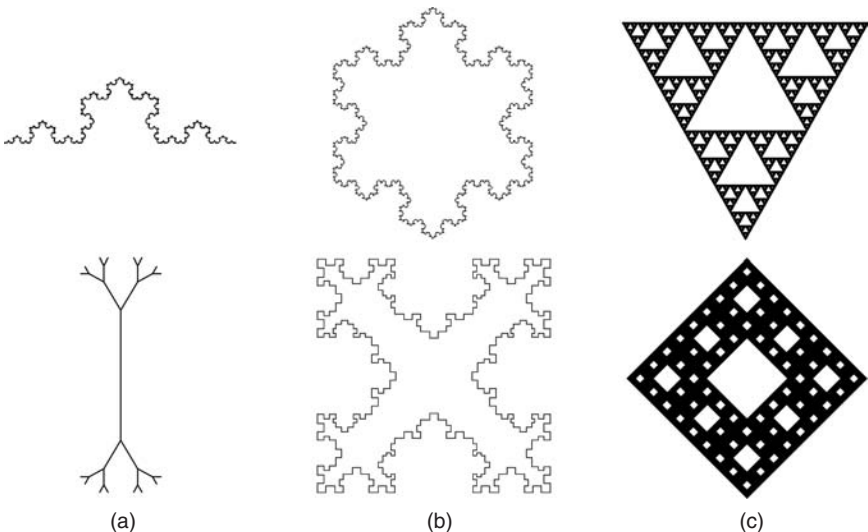


FIGURE 33-14 Some fractal geometries commonly used in antenna applications: (a) Koch curves and fractal trees have been used to create miniaturized dipole and monopole antenna elements. (b) Koch snowflakes and islands have been used to create loop and patch antennas. (c) Sierpinski gaskets and carpets have been used to create multiband antenna elements.

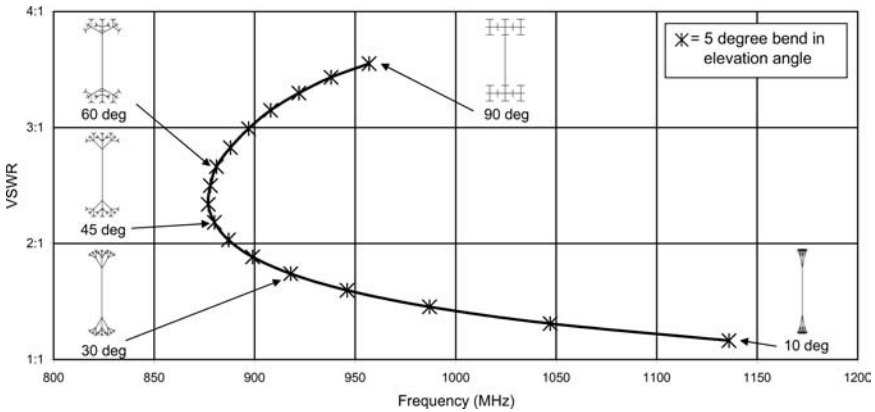


FIGURE 33-15 VSWR vs. elevation angle resonant frequency for a set of stage 3, four-branch fractal tree antennas with elevation angles ranging from 10° to 90°. The VSWR was calculated with respect to a 50- Ω line (after Petko and Werner¹⁰ © IEEE 2004).

resonant frequency than those with large elevation angles. This trend continues until the resonant frequency reaches a minimum for designs with elevation angles of approximately 50°. Designs with elevation angles beyond 50° are not optimal because both the resonant frequency and VSWR increase with the increase in elevation angle.

These trends have been used to design more effective miniature fractal tree dipoles, such as one created from a six-branch generator with branch elevation angles alternating between 30° and 50°. Figure 33-16 illustrates the geometry of the end-loads for the first three iterations of this dense fractal tree antenna. In Figure 33-17, the reflection properties (i.e., S_{11} and VSWR) of these dense fractal tree antennas are compared with the reflection properties of the four-branch fractal trees illustrated in Figure 33-5. The second-stage, six-branch 50°–30° antenna is shown to have a resonant frequency of 920 MHz, which is identical to the resonant frequency of the third iteration for the four-branch fractal tree antenna. The resonant frequency for the third-iteration, six-branch 50°–30° antenna is shown to be 790 MHz. This resonance occurs at a frequency 57 percent lower than the 1820-MHz resonance of an equivalently sized conventional Euclidian dipole antenna, and is 70 MHz lower than the fourth iteration of the four-branch fractal tree antenna.

One of the key advantages of end-loading a dipole antenna with a fractal-tree geometry is evident when LC resonators or RF switches are incorporated throughout the end-load structure. Resonators and switches are often used to make conventional antennas either multiband or reconfigurable; however, in conventional structures, it is difficult to create

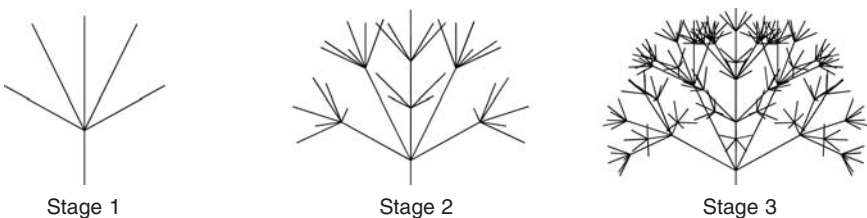


FIGURE 33-16 Fractal tree end-loads for the three-dimensional six-branch 30°–50° fractal tree dipole antenna (after Petko and Werner¹⁰ © IEEE 2004)

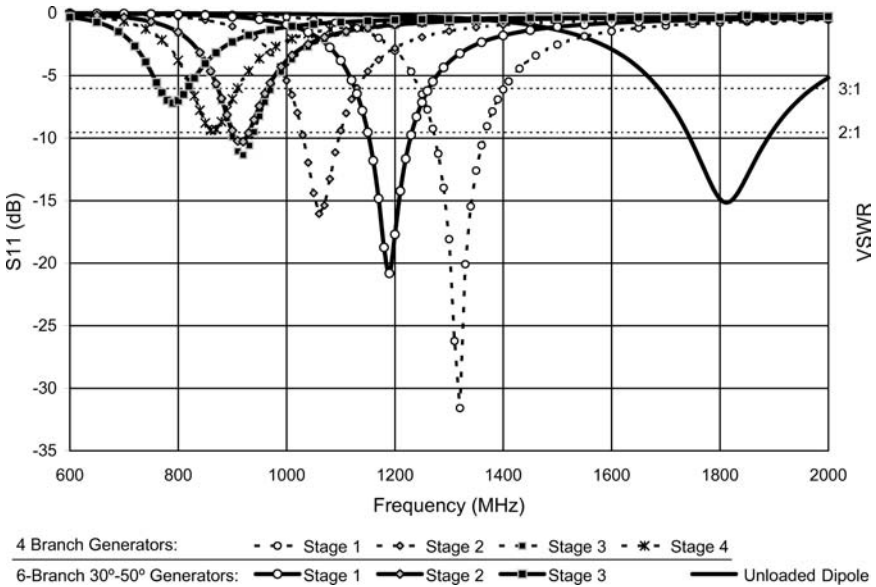


FIGURE 33-17 S_{11} versus frequency for a stage 1, stage 2, and stage 3 three-dimensional, six-branch 50°–30° fractal tree antenna. Also shown for comparison is the S_{11} for the first four stages of a standard four-branch fractal tree antenna and an equivalently sized conventional dipole. The S_{11} was calculated with respect to a 50- Ω line (after Petko and Werner¹⁰ © IEEE 2004).

resonances that are close to each other in frequency. This difficulty arises primarily because in order to achieve this type of behavior, the reactive loads and switches would have to be placed in close physical proximity to each other on the antenna. Moreover, the reactive loads or switches are usually distributed in a series fashion along the antenna, which can compound the losses associated with the use of these devices. In a fractal tree end-load, however, these problems do not exist because reactive loads and switches can be spread throughout the fractal structure, placing the loads into a parallel configuration and separating the required physical distance between them. These reactive loads and switches essentially change the configuration, and hence the size, of the fractal tree end-load for specific frequencies.

In one study, RF switches are strategically placed throughout the end-load of a third-stage, six-branch 50°–30° fractal tree dipole antenna (see Figure 33-16) to make it reconfigurable (i.e., tunable) over a bandwidth of 68 percent. The design uses 102 individual switches placed in each end-load of the antenna, as shown in Figure 33-18, to produce 20 reconfigurable states. The resulting dipole antenna was found to be reconfigurable from 770 MHz to 1570 MHz for a bandwidth of 800 MHz with a VSWR under 3:1 and reconfigurable from 970 MHz to 1570 MHz for a bandwidth of 560 MHz with a VSWR below 2:1. Also, since the design uses the fractal tree end-loads, the lowest resonance of this antenna occurs at a frequency 57 percent below that of a conventional linear dipole of equivalent length.

In Figure 33-19 each of the 20 reconfigurable states is represented by a separate S_{11} curve (indicated by light-gray lines) with the lowest resonant frequency representing the state with all the switches closed and the highest resonant frequency representing the state

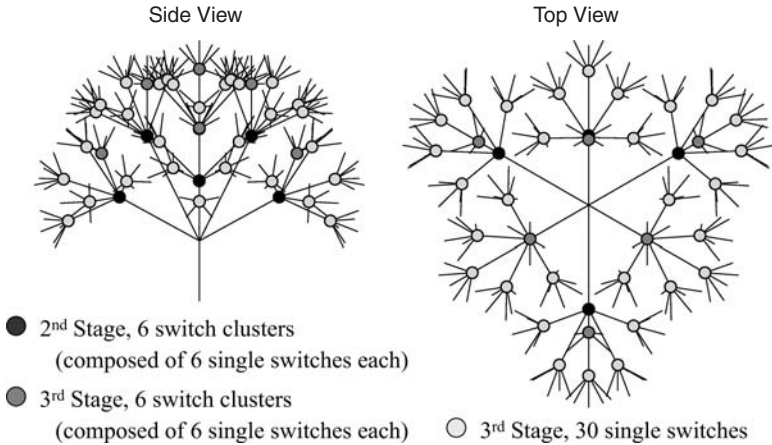


FIGURE 33-18 Switch layout for the reconfigurable six-branch, 50°–30° fractal tree dipole antenna (after Petko and Werner¹⁰ © IEEE 2004)

with all the switches open. The remaining states are achieved by opening the switches progressively from the top to the bottom. In addition, for three of the reconfigurable states the antenna effectively operates as a 50°–30° fractal tree dipole with fractal stages 3, 2, or 1. The S_{11} curves for these three special cases are indicated on the graph by thick dark-gray lines. Finally, the solid-black line represents the overall minimum S_{11} the antenna can be configured to for a particular frequency over the entire operating range of the antenna.

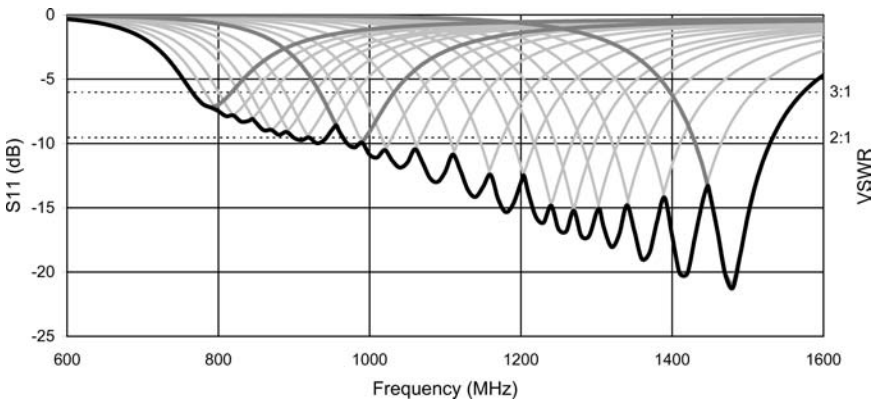


FIGURE 33-19 S_{11} versus frequency for the reconfigurable six-branch, 50°–30° fractal tree antenna. The light-gray curves represent each of the 20 states the antenna can be configured to operate. The dark-gray curves represent reconfigured states that operate as stage 1, stage 2, and stage 3 fractal tree antennas. The black line represents the overall minimum S_{11} the antenna can operate over the entire band. The S_{11} was calculated with respect to a 50- Ω line (after Petko and Werner¹⁰ © IEEE 2004)

33.5 FRACTAL ANTENNA ARRAYS

The advantages of fractal geometries are most evident when applied to antenna arrays. Their nature-based structure provides interesting radiation properties that are not found in conventional arrays. The radiation characteristics of deterministic fractal array configurations, such as Cantor linear arrays, Cantor ring arrays, and Sierpinski carpet planar arrays, have been considered in the literature.^{6,7,9,28,29} In some cases, these arrays were found to possess lower sidelobe levels than their conventional Euclidean counterparts. In addition, the fact that these fractal-based array geometries can be generated in an iterative fashion (i.e., via successive stages of growth starting from a simple generating array) has been exploited in order to develop rapid algorithms for use in efficient radiation pattern computations.^{6,7,9} Fractal concepts were also employed in Puente and Pous²⁸ to design linear arrays that exhibit multiband radiation characteristics. Moreover, a fractal radiation pattern synthesis technique has been developed in Werner et al³⁰ for the design of reconfigurable multiband linear and planar arrays.

In addition to deterministic fractal arrays, properties of random fractals have been used to develop a design methodology for quasi-random arrays, called fractal-random arrays.^{7,9,31} Inspiration for fractal-random geometries are found in the natural world, where many objects closely resemble fractals but are also not completely deterministic. Figure 33-4 illustrates the process of introducing randomness into a fractal-random object. Each time a generator is applied to construct a fractal structure, it is randomly selected from a set of multiple generators. In this manner, random fractals are used to generate array configurations that are somewhere between completely ordered (i.e., periodic) and completely disordered (i.e., random). An example of a fractal-random array layout is shown in Figure 33-20a. The main advantage of this technique is that it yields sparse arrays that possess relatively low sidelobe levels, which is a property indicative of periodic arrays, but over a range of bandwidths that are comparable to random arrays.

More recently, a robust and versatile design methodology for large size arrays has been developed that is based in part on the concept of an “optimal” fractal-random array.² This technique incorporates a genetic algorithm with a specialized subset of fractal-random arrays called polyfractal arrays. A genetic algorithm is a nature-based design tool that optimizes a problem globally using the Darwinian notions of natural selection and survival

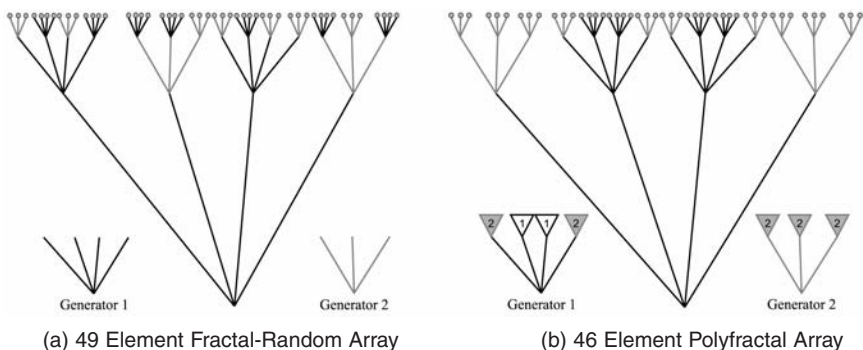


FIGURE 33-20 Fractal-random tree analogy for the construction of linear fractal-random and polyfractal arrays. The connection factors in (b) are illustrated by the numbers above the branches of the four-branch and the three-branch generators. Antenna elements are positioned at the ends of the topmost branches (after Petko and Werner² © IEEE 2005)

of the fittest. Useful overviews on the topic of genetic algorithms can be found in the literature.^{32,33,34} Although easily visualized, fractal-random geometries cannot exactly be reproduced from a small set of parameters and are not truly recursive when applied in an IFS. The subset of polyfractal arrays overcomes these limitations while retaining many of the desirable properties of fractal-random arrays.

To construct a polyfractal array, the IFS technique introduced in Section 33.3 must first be modified to handle multiple generators. Polyfractal arrays are constructed from multiple generators, $1, 2, \dots, M$, each of which has a corresponding Hutchinson operator W_1, W_2, \dots, W_M . Each Hutchinson operator W_m in turn contains N_m affine linear transformations, $\omega_{m,1}, \omega_{m,2}, \dots, \omega_{m,N_m}$. These affine linear transformations, $\omega_{m,n}$, are simplified from the more general six-parameter affine linear transformations introduced in Section 33.3 to a similitude approach consisting of three local parameters, $r_{m,n}$, $\varphi_{m,n}$, $\psi_{m,n}$, and one global-scale parameter s_f that is applied throughout the polyfractal structure. This affine linear transformation can be represented by

$$\omega_{m,n} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} s_f \cos(\varphi_{m,n} + \psi_{m,n}) & -s_f \sin(\varphi_{m,n} + \psi_{m,n}) \\ s_f \sin(\varphi_{m,n} + \psi_{m,n}) & s_f \cos(\varphi_{m,n} + \psi_{m,n}) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} r_{m,n} \cos \varphi_{m,n} \\ r_{m,n} \sin \varphi_{m,n} \end{pmatrix} \quad (33-7)$$

In addition to these three local parameters, a fourth local parameter, $\kappa_{m,n}$, is associated with each affine linear transformation. This parameter, referred to as the connection factor, is an integer value ranging from 1 to M , the number of generators used to construct the polyfractal array. In this generalized iterated function system approach, a Hutchinson operator, W_m , is used to construct a stage $\ell + 1$ polyfractal array from the set of possible stage ℓ polyfractal arrays, F_ℓ . Each affine linear transformation, $\omega_{m,n}$, can only be performed on stage ℓ polyfractal arrays where the generator applied at stage ℓ matches the connection factor, $\kappa_{m,n}$. Because the connection factors dictate how the affine linear transformations are applied, only one unique polyfractal array geometry can be associated with each Hutchinson operator. Therefore, the set of stage ℓ polyfractal arrays, F_ℓ , can be expressed by the notation

$$F_\ell = \{F_{\ell,1}, F_{\ell,2}, \dots, F_{\ell,M}\} \quad (33-8)$$

where the first subscript defines the level of the polyfractal array and the second subscript defines the generator employed at that level. Therefore, a polyfractal array of stage $\ell + 1$ constructed by generator m can be represented by

$$F_{\ell+1,m} = W_m(\{F_{\ell,1}, F_{\ell,2}, \dots, F_{\ell,M}\}) = \bigcup_{n=1}^{N_m} \omega_{m,n}(F_{\ell,\kappa_{m,n}}) \quad (33-9)$$

After the polyfractal structure is formed, the overall structure is adjusted by another global-scale parameter s_g . The global-scale factor s_f can be factored out of the Hutchinson operators to create the final normalized IFS construction procedure for a stage L polyfractal array:

$$F_{\ell,m}^L = s_g (s_f)^{L-\ell} \overline{W_m}(\{F_{\ell-1,1}^L, F_{\ell-1,2}^L, \dots, F_{\ell-1,M}^L\}) = s_g (s_f)^{L-\ell} \bigcup_{n=1}^{N_m} \overline{\omega_{m,n}}(F_{\ell-1,\kappa_{m,n}}^L) \quad (33-10)$$

This process is illustrated for a linear polyfractal array in Figure 33-20b using the fractal-tree analogy. In this figure, numbers representing the connection factors are attached to the ends of the generator branches. The branches ending with number 1 have generator 1 connected to their ends. Likewise, the branches ending with number 2 have generator 2 connected to their ends. The ends of the topmost branches represent the positions of the antenna elements.

Recursive beamforming algorithms can be created for both fractal and polyfractal arrays. Many fractal array recursive beamforming algorithms are based on the pattern multiplication approach. The radiation pattern of a stage ℓ fractal array is equal to the product of the radiation pattern of a stage $\ell-1$ fractal subarray and the array factor of the appropriately scaled fractal generator. In other words, the stage ℓ fractal array can be thought of as an array of stage $\ell-1$ fractal subarrays. An equation expressing the stage ℓ fractal subarray radiation pattern, FR_{ℓ}^L , is derived using the fractal similitude ω_n . The similitude is based on Eq. 33-7, using the local parameters, r_n , φ_n , and ψ_n and scale parameters s_g and s_f (the m is omitted because there is only one generator). To perform pattern multiplication, all subarray radiation patterns must be identical and oriented in the same direction. Therefore, the sum of φ_n and ψ_n is required to be equal to a multiple of 2π , making the axes of symmetry of the subarrays parallel. The equation for a recursive beamforming algorithm based on pattern multiplication can be written as

$$FR_{\ell}^L(\theta, \varphi) = FR_{\ell-1}^L(\theta, \varphi) \sum_{n=1}^N \exp j \left[k \left(s_g (s_f)^{L-\ell} r_n \right) \sin \theta \cos(\varphi - \varphi_n) \right] \quad (33-11)$$

Using isotropic sources as the initial subarray radiation pattern, the final stage L fractal array factor can be written in a similar manner as has been done in Werner et al⁶:

$$AF_L^L(\theta, \varphi) = \prod_{\ell=1}^L \sum_{n=1}^N \exp j \left[k S \delta^{\ell-1} r_n \sin \theta \cos(\varphi - \varphi_n) \right] \quad (33-12)$$

where $\delta = 1/s_f$ and $S = s_g (s_f)^{L-1}$. Typically, the values of r_n are scaled such that S can be set equal to one.

The unique scaling procedure and connection factor-based construction allow the rapid recursive beamforming algorithms associated with fractal arrays to be generalized to handle polyfractal arrays. The fractal array recursive beamforming operation discussed above requires all subarrays to have the same radiation pattern and be oriented in the same direction. In that way, pattern multiplication can be employed. In the more general polyfractal array, there are multiple types of subarrays that do not necessarily point in the same direction. Therefore, these subarray patterns cannot be factored out of the sum and the resulting expression for the stage ℓ , generator m subarray pattern is given by

$$FR_{\ell,m}^L(\theta, \varphi) = \sum_{n=1}^{N_m} \left(FR_{\ell-1,k_{m,n}}^L(\theta, \varphi - \varphi_{m,n} - \psi_{m,n}) \right) \exp j \left[k \left(s_g (s_f)^{L-\ell} r_{m,n} \right) \sin \theta \cos(\varphi - \varphi_{m,n}) \right] \quad (33-13)$$

This subarray radiation pattern is based on the set of stage $\ell-1$ fractal subarray patterns. The final radiation pattern can be determined by using isotropic sources for the initial subarray radiation patterns and recursively applying the expression until the stage L radiation pattern is obtained. Figure 33-21 illustrates this process for the example array shown in Figure 33-20b. One of the main advantages of the recursive beamforming approaches associated with fractal and polyfractal arrays is that they can be exploited to considerably speed up the convergence of the genetic algorithm (GA). This allows the possibility of optimizing much larger size arrays than has previously been possible using other approaches.

While conventional genetic algorithm techniques can be applied to polyfractal arrays of the same size, it is advantageous to generalize the optimization process to include polyfractal arrays of varying sizes. To accomplish this, the genetic algorithm crossover routine, where parameters of two parent arrays are combined to create new offspring, must be modified so that the parent chromosomes of the polyfractal arrays are broken apart into simpler pieces. This process, discussed in more detail in Petko and Werner,² essentially combines

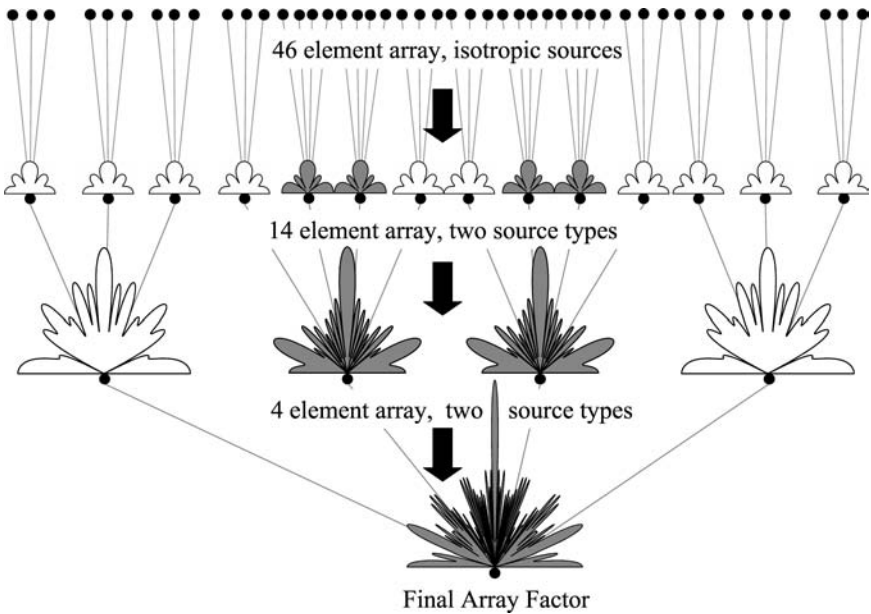


FIGURE 33-21 Rapid recursive beamforming algorithm for a two-generator polyfractal array (after Petko and Werner² © IEEE 2005)

various affine linear transformations from each parent array and performs crossover at that level. The resulting transformations are then combined into the offspring arrays. The genetic algorithm mutation routine can also be generalized for polyfractal arrays of varying sizes. Genetic mutation typically consists of changing a single parameter; however, polyfractal arrays offer a wide range of flexibility. Mutations can be performed on any of the global parameters as well as on any transformation. In addition, mutation can also remove or add a gene to a generator. Mutation can also be used to switch the order of the generators. This process is also discussed in more detail in Petko and Werner.² By allowing the algorithm to handle more general polyfractal array geometries, an added degree of flexibility is incorporated into the optimization process.

When the optimization appears to reach premature convergence for a population of polyfractal arrays, a unique mutation process, called *generator autopolyloidization*, can be used to stimulate the evolution process.³⁵ The generator autopolyloidization process divides each fractal-random generator into two identical parts. The connection factors that were used to select the previous generator are uniformly divided to choose between the two new generators. In this way the arrays are exactly the same; however, they are described using twice the number of local parameters, adding new flexibility for genetic evolution. After the generator duplication, each chromosome parameter is mutated a small amount through a perturbation process. This perturbation adds a degree of genetic diversity to the population that can aid in the overall evolution procedure. Finally, the population is ready for another period, or epoch, of genetic evolution. An example of this generator autopolyoid process is illustrated in Figure 33-22 for the transformation of a one-generator fractal array into a two-generator polyfractal array.

It is important to note how generator autopolyloidization can benefit the optimization process. Initially, the genetic algorithm is used to evolve simple solutions consisting of a small number of generators. The limited number of generators restricts the initial antenna

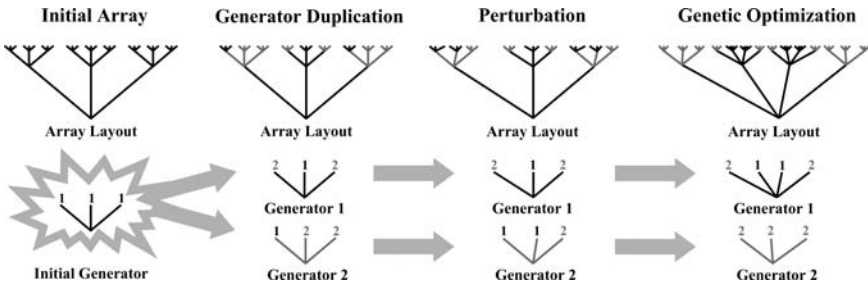


FIGURE 33-22 Generator autopolyloidization process for the conversion of a one-generator fractal array to a two-generator polyfractal array (after Petko and Werner³⁵ © IEEE 2007)

geometries to a small area of the search space; however, the time required to evaluate the performance of each antenna is very small. The optimization takes advantage of this short evaluation time until convergence is reached. At this point the algorithm can get no further using this limited set of generators; therefore, the generator autopolyloidization process is applied to expand the number of generators, which in turn expands the search space. This expansion comes at the cost of increasing the evaluation time; however, the optimization now has a good idea of where to search for more complex solutions based on the simpler designs found during the previous epoch. On the contrary, if the genetic algorithm began by optimizing the more complex solutions, not only would it initially take much longer to evaluate each array, the large number of unfocused input parameters could overwhelm the optimization process. This cycle of autopolyloidization can be repeated numerous times, producing more and more complex solutions with each optimization epoch. If the evolutionary progress is summarized for each generation, the resulting diagram would exhibit a staircase pattern. Thus, generator autopolyloidization offers polyfractal-based genetic algorithms an efficient “path” to complex design solutions, first optimizing simple designs very quickly and then only adding the increasing levels of complexity when they are finally needed.

Next, an example is presented to illustrate the GA design methodology for evolving an optimal uniformly excited linear polyfractal array. The design considered here is an example of a large- N array, i.e., an array that has a relatively large number of elements. In this case, an initial population of arrays is optimized to minimize the peak sidelobe level and maintain a narrow beamwidth for a polyfractal array with a 0.5λ minimum interelement spacing. The initial population was based on 500, 2401-element periodic arrays with 0.5λ spacing. These initial arrays were created using a polyfractal structure consisting of one generator. Three generator autopolyloidy operations were performed: one initially and two triggered after the population did not improve over 30 generations. Therefore the optimization process is divided into three epochs of evolution. The evolutionary diagram, shown in Figure 33-23, exhibits the characteristic staircase pattern resulting from the three generator autopolyloidy operations. During the first epoch, the polyfractal arrays had two generators. This epoch ended at generation 217, where the number of generators was doubled to four for the second epoch. Two hundred generations later, the third epoch began with each polyfractal array now having eight generators. The final 1616-element design was found after 700 generations and had a sidelobe level of -24.30 dB and a beamwidth of 0.056° . The recursive beamforming algorithm on average calculated radiation patterns 20 times faster for arrays with two generators, 15 times faster for arrays with four generators, and 10 times faster for arrays with eight generators when compared to a conventional discrete Fourier transform (DFT) approach. Table 33-1 summarizes these speed increases for each epoch. The array factor and geometrical layout of this antenna array are shown in Figure 33-24 with the corresponding performance properties of the antenna listed in Table 33-2.

TABLE 33-1 Number of Generators and Radiation Pattern Evaluation Speed Increase per Epoch for the 1616-element Polyfractal Array Using the Recursive Beamforming Algorithm vs. a Conventional DFT Approach (Radiation patterns have been sampled at 72,000 points.)

	Epoch 1	Epoch 2	Epoch 3
Number of Generators	2	4	8
Evaluation Speed Increase	1937%	1431%	892%

TABLE 33-2 Performance Properties of a 1616-element Genetically Optimized Polyfractal Array

Number of Elements	SLL (dB)	HPBW	Minimum Spacing	Average Spacing
1616	-24.30	0.056°	0.5λ	1.01λ

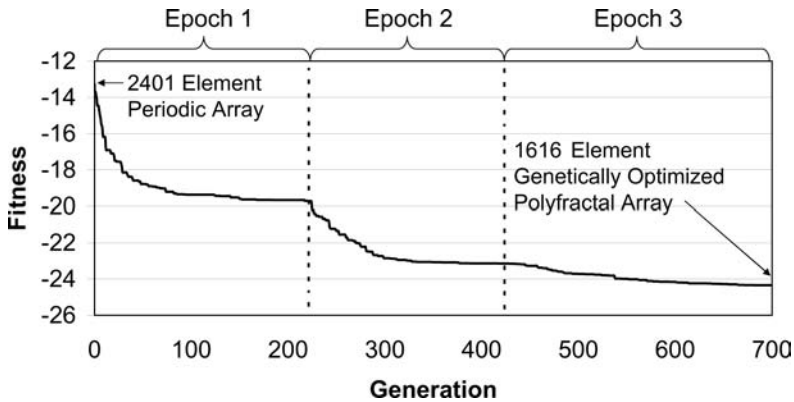


FIGURE 33-23 Evolutionary diagram of a 1616-element genetically optimized polyfractal array (after Petko and Werner³⁵ © IEEE 2007)

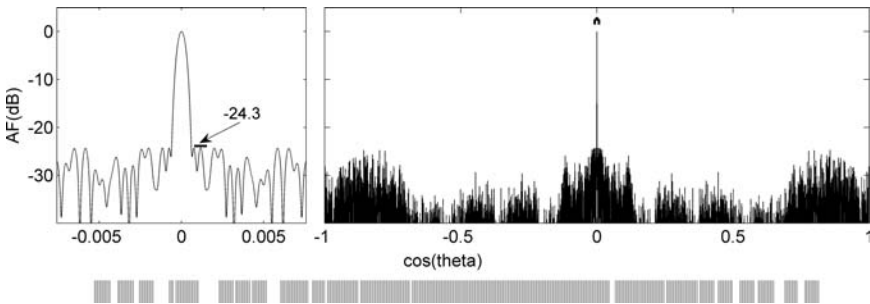


FIGURE 33-24 Radiation pattern and array layout for a 1616-element genetically optimized polyfractal array (after Petko and Werner³⁵ © IEEE 2007)



FIGURE 33-25 Layout for a 1230-element, uniformly excited, Pareto optimized polyfractal array.

More complex design problems can have multiple criteria that must be optimized. In these cases, it is desirable to compare the solutions that lie upon the Pareto front (i.e., the set of all nondominated solutions). Several different types of genetic algorithms have been developed to search for this Pareto front.^{36–43} An example of a uniformly excited polyfractal array evolved from one of these Pareto genetic algorithm optimization techniques^{39,40} is discussed here. The goal of the optimization is to minimize the peak sidelobe level at two frequencies: one in which the minimum spacing is 0.75λ and another where the minimum spacing is 2.0λ . The optimization evolved arrays from an initial population of fractal arrays based on a 1296-element periodic array. Three generator autopolyploidy processes were performed: one initially, one at generation 150, and one at generation 300. The optimization was terminated after 450 generations and the resulting Pareto front consists of arrays with between 1186 and 1230 elements. One solution with 1230 elements was chosen from the Pareto front for discussion. This example has a -19.3 -dB sidelobe level at both the 0.75λ and at the 2.0λ minimum spacings when the mainbeam is pointed at broadside. The layout of this design is shown in Figure 33-25, while Table 33-3 summarizes the properties of this array at the two different design frequencies.

While the array was optimized only at these two frequencies, the fractal-random properties of polyfractal arrays allow for low sidelobe levels over fairly wide ranges of bandwidth without the appearance of grating lobes. Figure 33-26 plots the sidelobe level performance of the polyfractal array shown in Figure 33-25 as a function of the minimum separation between elements in wavelengths and compares them to a 1296-element periodic array. When steered to broadside, the sidelobe level of the periodic array is -13.26 dB until the electrical spacing between elements increases to about 1λ , when grating lobes first appear. In contrast, the polyfractal array maintains sidelobe levels less than -14 dB even up to 20λ minimum spacings. This array was not optimized at a 20λ minimum spacing; nevertheless, the robust physical properties of the polyfractal arrays in combination with the optimization goals at 0.75λ and 2.0λ minimum spacings have allowed the low sidelobe performance to be maintained far beyond the specified separation distance of 2.0λ . This array also has a sidelobe level under -16 dB for up to 16λ minimum spacing, and a sidelobe level under -17 dB up to 7λ minimum spacing when the mainbeam is directed to broadside. In addition, one can see how the goals of the optimizer at the 0.75λ and 2.0λ minimum spacings have evolved an array with sidelobe levels under -19 dB for these lower frequencies.

TABLE 33-3 Performance at the Two Optimization Target Frequencies for the 1230-element Polyfractal Array

Number of Elements	0.75 λ Minimum Spacing			2.0 λ Minimum Spacing		
	<i>SLL</i>	<i>HPBW</i>	<i>Average Spacing</i>	<i>SLL</i>	<i>HPBW</i>	<i>Average Spacing</i>
1230	-19.31 dB	0.0078°	5.60λ	-19.31 dB	0.0029°	14.96λ

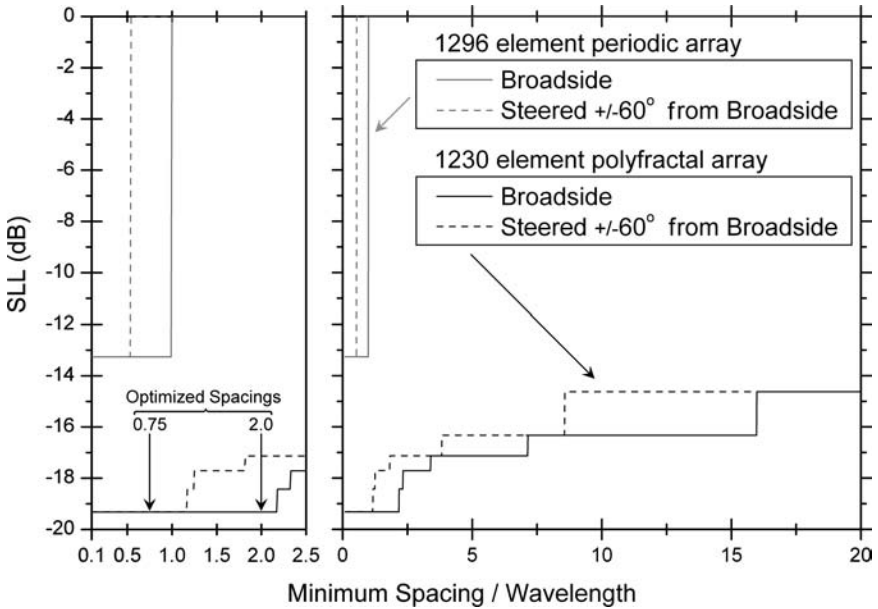


FIGURE 33-26 Sidelobe level performance of a uniformly excited, 1230-element, Pareto optimized polyfractal array over a range of bandwidth compared to the performance of a conventional uniformly excited periodic linear array. The polyfractal array was optimized at broadside for simultaneous minimum element spacings of 0.75λ and 2.0λ .

Finally, these arrays also exhibit their wideband properties when they are steered away from broadside. Figure 33-26 also plots the peak sidelobe level vs. minimum spacing when the mainbeam is steered 60 degrees from broadside. The grating lobes for the periodic array when steered to 60 degrees from broadside first appear at interelement spacings of 0.56λ . On the other hand, the peak sidelobe level performance of the polyfractal array is still less than -14 dB even up to 20λ minimum spacing. Figure 33-27 shows the radiation pattern for this optimized polyfractal array when the minimum spacing between elements is 8λ and the beam is steered 60 degrees from broadside.

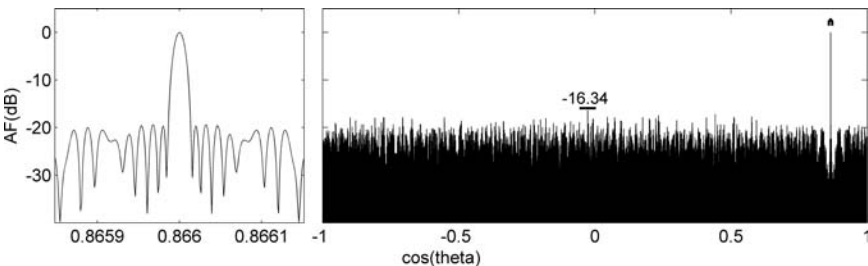


FIGURE 33-27 Radiation pattern for a uniformly excited, 1230-element, Pareto optimized polyfractal array with its mainbeam steered 60 degrees from broadside and an 8λ minimum spacing between elements

33.6 ANTENNA ARRAYS BASED ON FRACTAL AND APERIODIC TILINGS

This section is devoted to design methodologies for antenna arrays that are based on fractal and aperiodic tilings. Although a wide variety of tilings ranging from simple to complex patterns have been known since early civilizations (an example of an early tiling is shown in Figure 33-28), the formal mathematical theory of tilings has only been in existence for about a century.⁴⁴ Tiling theory has found applications in many fields, such as crystallography, biology, and communication theory. More recently, tiling theory has been successfully applied to the field of electromagnetics, in particular, antenna array design using fractal tilings^{45,46,47} and aperiodic tilings.⁴⁸ The unique geometry of these tilings can be exploited to generate antenna arrays that exhibit low sidelobe levels and suppressed grating lobes for minimum element spacings of at least one-wavelength. Moreover, it has been shown that robust element perturbation schemes based on the GA can be used to greatly improve the performance of these arrays.^{49,50,51}

A modular design approach based on the theory of fractile geometry has been introduced in Werner et al.⁴⁵ for designing low-sidelobe arrays with suppressed grating lobes. A fractile array is defined as an array composed of a tiling of self-similar subarrays with fractal boundaries that covers the plane or a portion of the plane without any overlaps or gaps. A variety of different fractile array configurations have been investigated including Peano-Gosper, terdragon, six-terdragon, and fudgeflakes.⁴⁵ Another important property of fractile arrays is that their self-similar tile geometry can be exploited to develop a rapid iterative procedure for calculating far-field radiation patterns that, for sufficiently large arrays, can be considerably faster than using a conventional DFT.

One specific type of fractile array that has been reported on extensively in the literature is based on the Peano-Gosper family of space-filling curves (see Figure 33-7) and is therefore known as the Peano-Gosper fractile array (PGFA).^{45,46,47,49,50} The elements of the array are uniformly distributed along a Peano-Gosper curve as shown in Figure 33-29, which leads to a planar array configuration with an equilateral triangular (i.e., hexagonal) lattice on the interior that is bounded by an irregular closed Koch fractal curve around its perimeter. The Koch fractal boundary plus its interior form a Gosper island that can be used to cover the plane

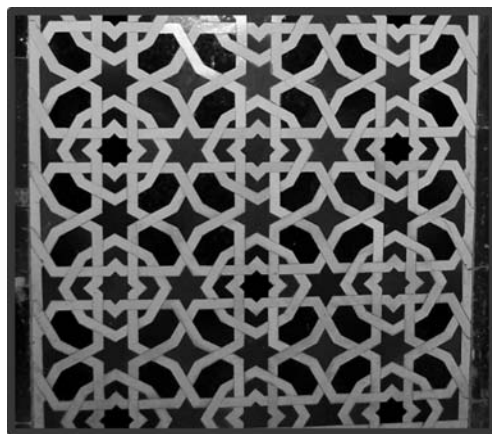


FIGURE 33-28 Photograph of a tiling from the Alhambra in Granada, Spain (Courtesy of Douglas H. Werner)

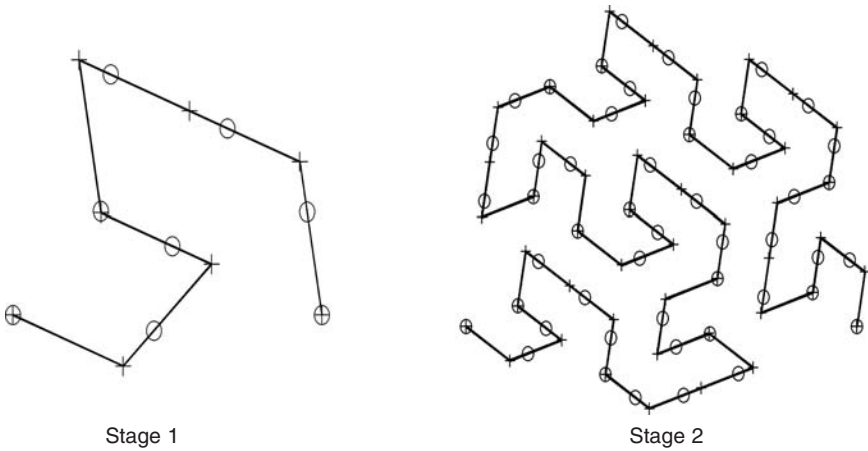


FIGURE 33-29 Element distribution along stage 1 and stage 2 of the Peano-Gosper curve. The initial uniformly spaced element locations are denoted by a +. Examples of GA perturbed element locations are denoted by an o.

via a tiling.^{45,46} Furthermore, Gosper island tiles as illustrated in Figure 33-9 are self-similar and can be divided into seven smaller tiles, each representing a scaled copy of the original. Finally, the PGFA may be iteratively constructed to any arbitrary stage of growth based on a set formula for shifting, scaling, and rotating the generating array defined at stage 1. These unique properties were exploited in Werner et al⁴⁵ and Bogard et al⁴⁷ to develop a design methodology for deterministic arrays that possess several desirable properties, including a uniform current distribution, low sidelobe levels, relatively broad bandwidth, modular architecture, and the ability to perform rapid beamforming. Moreover, due to its equilateral triangular lattice, the PGFA will have no grating lobes for broadside operation even when the minimum spacing between elements is increased to one wavelength.

Because the PGFA possesses a fractal boundary contour, its sidelobe levels are lower than an equivalent size rectangular array with an equilateral triangular (i.e., hexagonal) grid of elements on its interior. Figure 33-30 shows a comparison of the array factor for a 352-element rectangular-hexagonal array, a 344-element stage 3 PGFA, and a 361-element (19×19) square periodic array.⁴⁷ All of the arrays have uniform element excitation and a minimum element spacing of one-wavelength. At this spacing the PGFA exhibits lower overall sidelobes than either its conventional square periodic or rectangular-hexagonal array counterparts.

In the design of phased arrays (especially broadband phased arrays), care must be taken to ensure that grating lobes do not appear when the main beam is to be steered away from broadside. For example, when the element spacing of a PGFA is increased to one wavelength, no grating lobes occur in the entire visible region for broadside operation. However, when the minimum element spacing is one wavelength or greater and the main beam is scanned away from broadside, grating lobes occur in the far-field radiation pattern of the array. It has been shown in Bogard and Werner⁴⁹ and in Bogard et al⁵⁰ that a GA technique can be used to perturb the element locations in an optimal way on the interior of the PGFA to eliminate the grating lobes and provide acceptable sidelobe levels during scanning. This procedure results in a modular or tiled phased-array architecture that has an irregular fractal distribution of elements around its periphery together with an aperiodic arrangement of elements on its interior. During the design process, the element locations of the stage 1 generator are varied and subsequently used to generate higher-order stages of the PGFA through an efficient iterative procedure while, at each stage, the array maintains its broadband

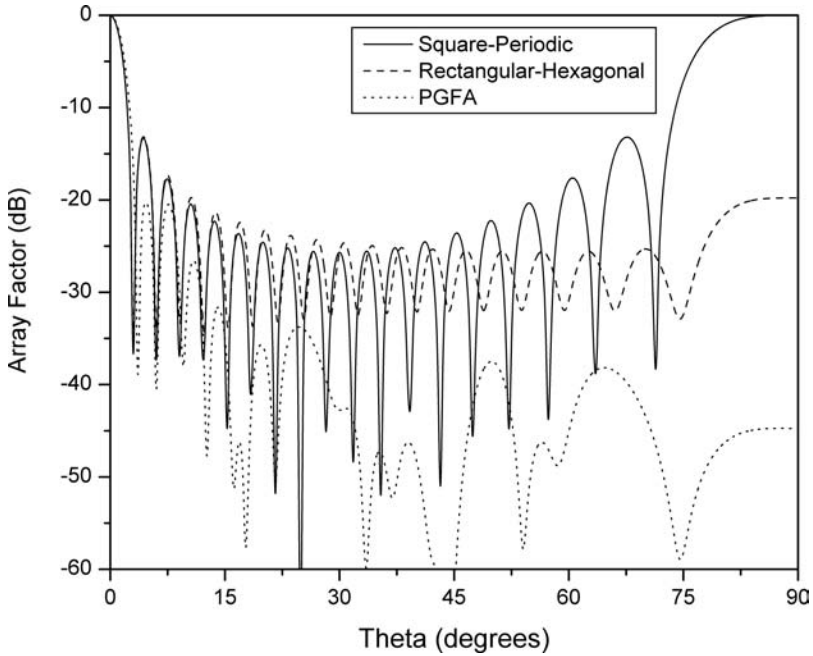


FIGURE 33-30 Comparison of the normalized array factor versus θ for a 361-element (19×19) square periodic array, a 352-element rectangular-hexagonal array, and a 344-element stage 3 PGFA at $\varphi = 0$. In all three cases the array elements are uniformly excited and they have a minimum element spacing of one wavelength.

characteristics within a specified scan volume. Examples of GA perturbed element locations along the Peano-Gosper curve are shown in Figure 33-29.

In Bogard and Werner⁴⁹ and in Bogard et al.,⁵⁰ a stage 3 PGFA was optimized and its performance was compared with that of a uniform PGFA. The array had an initial element spacing of 2λ and was optimized for scanning up to $\theta = 30^\circ$. During scanning, grating lobes are suppressed over the entire visible region of the GA optimized design. Figure 33-31

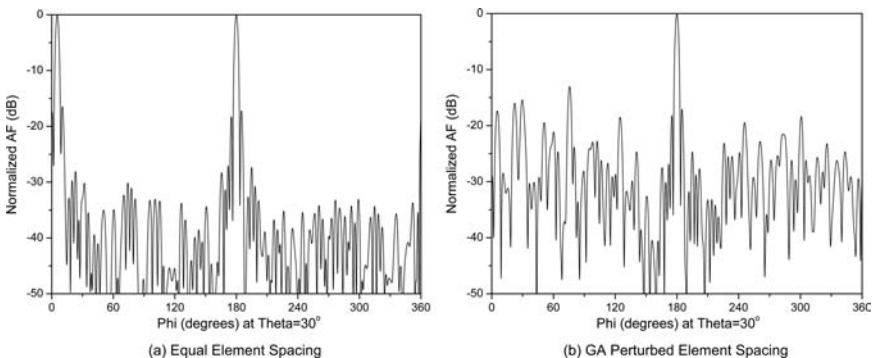


FIGURE 33-31 Normalized array factor cut for the stage 3 PGFA with an initial element spacing of 2λ . The main beam of the array was steered to $\theta = 30^\circ$, $\varphi = 180^\circ$.

shows a $\theta = 30^\circ$ cut of the array factor for the optimized PGFA and the uniform PGFA when the main beam is steered to $\theta = 30^\circ$ and $\varphi = 180^\circ$. In this plane cut, a grating lobe is present in the array factor of the original uniform PGFA, whereas the optimized PGFA has complete grating lobe suppression over the same scanning range.

Analogous to fractal-based tilings, aperiodic tilings can be used to generate antenna arrays that exhibit properties that are not typically associated with conventional periodic arrays. Unlike fractile arrays, which consist of a collection of elements that are placed a uniform distance apart along a space-filling curve (e.g. a Peano-Gosper curve), aperiodic tiling-based arrays consist of elements that are located at the vertices of an aperiodic tiling lattice.⁴⁸ Arrays that are generated in this manner tend to have geometrical structures that contain local order and rotational symmetry, but are devoid of any translational symmetries. A common method that is employed to generate these tilings is based on a decomposition process,⁴⁴ which is similar in some ways to the IFS techniques used to generate fractal geometries. In the process, tiles are decomposed into a collection of smaller tiles, which are either scaled copies of the original tiles or other tiles. This iterative process is continued until a large tiling is created.

The Danzer tiling is an example of an aperiodic tiling that can be used to generate broadband low-sidelobe planar antenna arrays that are naturally thinned.^{48,51} Moreover, the Danzer tiling is a type of aperiodic tiling that comprises a collection of three specific triangular-shaped prototiles. The tiling is formed by covering the plane with the triangles while maintaining specific matching rules throughout. A portion of a Danzer tiling is shown in Figure 33-32 along with its three prototiles.⁴⁴ An example of a Danzer tiling-based array is shown in Figure 33-33. The peak sidelobe level versus the minimum element spacing in terms of wavelengths for the array is shown in Figure 33-34. The tiling for the array has been scaled such that its physical dimensions correspond to a minimum element spacing, d , of $\lambda/2$ at the lowest intended operating frequency, and it has been truncated to have a circular aperture with a 12λ radius. In Figure 33-34, the performance of the Danzer array is compared with a conventional periodic array that has the same circular aperture and minimum element spacing. It is clear that the Danzer array outperforms the periodic array in terms of grating lobe suppression for large element spacings. Additionally, to fit within the same aperture size, the conventional periodic array requires approximately 1793 elements, while the naturally thinned (i.e., its mean interelement spacing is greater than $\lambda/2$) Danzer array only requires 811 elements.

Given the intrinsic properties of this class of arrays, they are attractive candidates for use as the baseline in various antenna designs, such as generating very sparse arrays⁵¹ and arrays with very wide bandwidth. One design methodology that is capable of producing

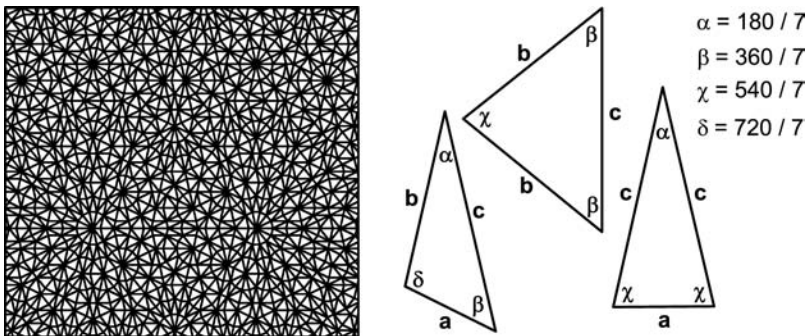


FIGURE 33-32 Truncated region of a Danzer tiling and its three prototiles

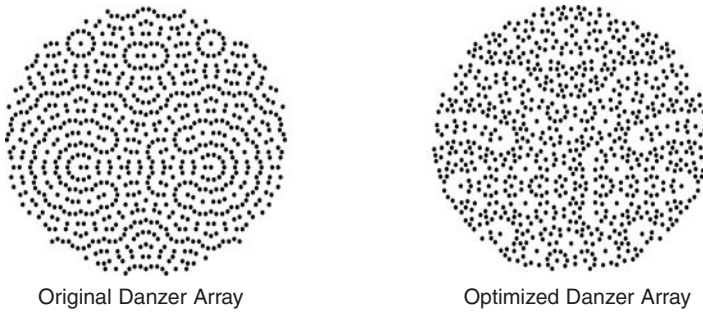


FIGURE 33-33 Geometry of the initial Danzer array and the GA optimized Danzer array

wideband planar arrays is based on a simple perturbation of the basic aperiodic array generation process. In this technique, an additional point (or points) is placed within the boundary of the prototiles that are used to generate the aperiodic tiling. The locations of these points are preserved within the prototiles as the overall aperiodic tiling is generated via the decomposition process. The result of this process is the formation of an aperiodic tiling that contains an additional point within each of its constituent tiles. Converting this to an antenna array yields fundamental elements at the vertices of the tiling along with elements at each of the additional points. The array can be scaled to have a specific minimum element spacing and then truncated to fit within a desired aperture. By adjusting the position of the point within each base tile, it is possible to greatly vary the radiation properties of the modified aperiodic tile-based array.

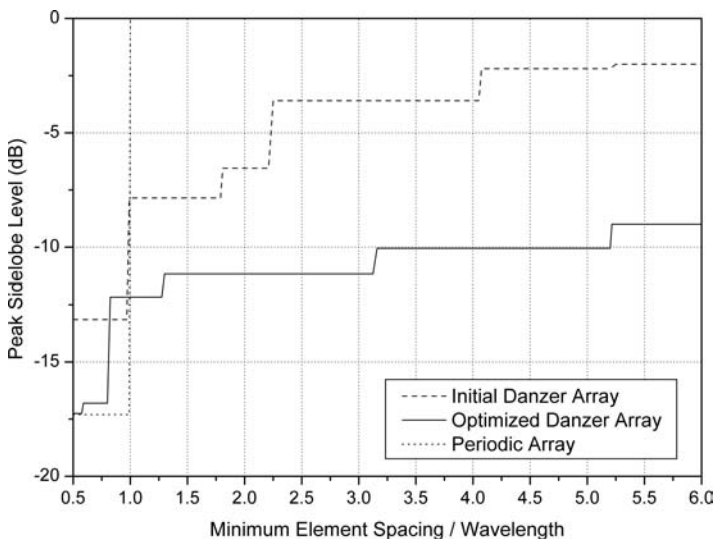


FIGURE 33-34 Sidelobe level performance of the initial Danzer array, GA optimized Danzer array, and periodic array. All of the arrays have a circular aperture with a radius of 12λ and a minimum element spacing of $d = 0.5\lambda$ at the lowest operating frequency.

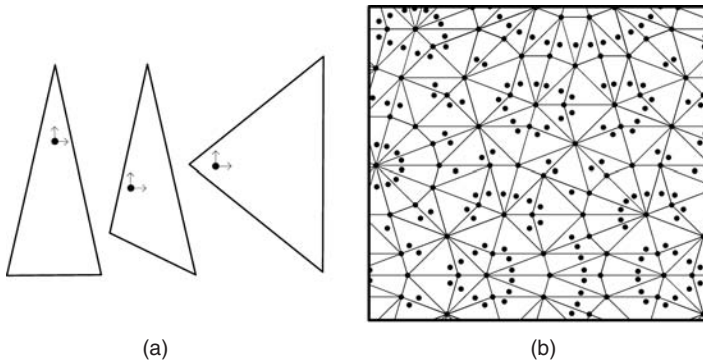


FIGURE 33-35 (a) Perturbation points of the Danzer prototiles and (b) an example of a truncated Danzer tiling with its corresponding perturbed array element locations

A GA-based optimizer was combined with the perturbation technique with the objective of designing a Danzer array that has the lowest possible sidelobe levels at $d/\lambda = 5$. Since the Danzer array comprises a collection of three prototiles, the GA was only required to optimize the planar coordinates of the three additional perturbation points, as illustrated in Figure 33-35. During the optimization, all of the generated arrays were scaled to have a minimum element spacing of $\lambda/2$ and were truncated to have a circular aperture with a radius of 12λ at the lowest operating frequency. The base tiling for the optimization was the same one that was used in the formation of the array shown in Figure 33-33. The design that results from this optimization process is shown in Figure 33-33 and has a normalized radiation pattern with a maximum sidelobe level of -10.05 dB at a frequency corresponding to $d/\lambda = 5$. The peak sidelobe level versus frequency for the array is shown in Figure 33-34. For this particular example, the simple perturbation scheme is capable of greatly extending the bandwidth (peak sidelobe level ≤ -10 dB with no grating lobes) of the basic Danzer array.

REFERENCES

1. K. M. Neiss, D. H. Werner, M. G. Bray, and S. Mummaready, "Nature-Based Antenna Design: Interpolating the Input Impedance of Fractal Dipole Antennas via a Genetic Algorithm Trained Neural Network," *Proc. USNC/URSI Nat. Radio Science Meeting* (June 2002): 374.
2. J. S. Petko and D. H. Werner, "The Evolution of Optimal Linear Polyfractal Arrays Using Genetic Algorithms," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 11 (November 2005): 3604–3615.
3. B. B. Mandelbrot, *The Fractal Geometry of Nature* (New York: W. H. Freeman, 1983).
4. H. O. Peitgen, H. Jurgens, and D. Saupe, *Chaos and Fractals: New Frontiers of Science* (New York: Springer-Verlag, Inc., 1992).
5. D. L. Jaggard, "Fractal Electrodynamics: From Super Antennas to Superlattices," *Fractals in Engineering*, J. L. Vehl, E. Lutton, and C. Tricot (eds.) (New York: Springer-Verlag, 1997).
6. D. H. Werner, R. L. Haupt, and P. L. Werner, "Fractal Antenna Engineering: The Theory and Design of Fractal Antenna Arrays," *IEEE Antennas and Propagation Magazine*, vol. 41, no. 5 (October 1999): 37–59.
7. D. H. Werner and R. Mittra (eds.), *Frontiers in Electromagnetics*, Chaps. 2 and 3 (Piscataway, New Jersey: IEEE Press, 2000).
8. J. P. Gianvittorio and Y. Rahmat-Samii, "Fractal Antennas: A Novel Antenna Miniaturization Technique, and Applications," *IEEE Antennas and Propagation Magazine*, vol. 44, no. 1 (February 2002): 20–36.

9. D. H. Werner and S. Ganguly, "An Overview of Fractal Antenna Engineering Research," *IEEE Antennas and Propagation Magazine*, vol. 45, no. 1 (February 2003): 38–57.
10. J. S. Petko and D. H. Werner, "Miniature Reconfigurable Three-Dimensional Fractal Tree Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 8 (August 2004): 1945–1956.
11. H. Sagan, *Space-Filling Curves* (New York: Springer-Verlag, 1994).
12. G. A. Edgar, *Measure, Topology, and Fractal Geometry* (New York: Springer-Verlag, 1990).
13. M. F. Barnsley, *Fractals Everywhere*, 2nd Ed. (New York: Academic Press Professional, 1993).
14. N. Cohen, "Fractal Antennas: Part 1," *Communications Quarterly* (Summer 1995): 7–22.
15. C. Borja and J. Romeu, "On the Behavior of Koch Island Fractal Boundary Microstrip Patch Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 6 (June 2003): 1281–1291.
16. T. G. Spence and D. H. Werner, "Genetically Optimized Fractile Microstrip Patch Antennas," *Proc. 2004 IEEE Antennas and Propagation International Symposium*, vol. IV (June 21–26, 2004, Monterey, CA): 4424–4427.
17. C. Puente, J. Romeu, R. Pous, X. Garcia, and F. Benitez, "Fractal Multiband Antenna Based on the Sierpinski Gasket," *IEEE Electronics Letters*, vol. 32, no. 1 (January 1996): 1–2.
18. C. Puente, J. Romeu, R. Pous, and A. Cardama, "On the Behavior of the Sierpinski Multiband Fractal Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 46, no. 4 (April 1998): 517–524.
19. G. J. Walker and J. R. James, "Fractal Volume Antennas," *IEEE Electronics Letters*, vol. 34, no. 16 (August 1998): 1536–1537.
20. C. T. P. Song, P. S. Hall, H. Ghafouri-Shiraz, and D. Wake, "Fractal Stacked Monopole with Very Wide Bandwidth," *IEEE Electronics Letters*, vol. 35, no. 12 (June 1999): 945–946.
21. N. Cohen, "Fractal and Shaped Dipoles," *Communications Quarterly* (Spring 1996): 25–36.
22. C. Puente, J. Romeu, R. Pous, J. Ramis, and A. Hijazo, "Small But Long Koch Fractal Monopole," *IEEE Electronics Letters*, vol. 34, no. 1 (January 1998): 9–10.
23. C. P. Baliarda, J. Romeu, and A. Cardama, "The Koch Monopole: A Small Fractal Antenna," *IEEE Transactions on Antennas and Propagation*, vol. 48, no. 11 (November 2000): 1773–1781.
24. C. Puente, J. Claret, F. Sagues, J. Romeu, M. Q. Lopez-Salvans, and R. Pous, "Multiband Properties of a Fractal Tree Antenna Generated by Electrochemical Deposition," *Electronics Letters*, vol. 32, no. 25 (December 1996): 2298–2299.
25. M. Sindou, G. Ablart, and C. Sourdis, "Multiband and Wideband Properties of Printed Fractal Branched Antennas," *Electronics Letters*, vol. 35, no. 3 (February 1999): 181–182.
26. D. H. Werner, A. Rubio Bretones, and B. R. Long, "Radiation Characteristics of Thin-wire Ternary Fractal Trees," *Electronics Letters*, vol. 35, no. 8 (April 1999): 609–610.
27. K. J. Vinoy, J. K. Abraham, and V. K. Varadan, "Fractal Dimension and Frequency Response of Fractal Shaped Antennas," *2003 IEEE Antennas and Propagation Society International Symposium*, vol. 4 (June 22–27, 2003): 222–225.
28. C. Puente and R. Pous, "Fractal Design of Multiband and Low Side-lobe Arrays," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 5 (May 1996): 730–739.
29. D. L. Jaggard and A. D. Jaggard, "Cantor Ring Arrays," *Microwave and Optical Technology Letters*, vol. 19 (1998): 121–125.
30. D. H. Werner, M. A. Gingrich, and P. L. Werner, "A Self-Similar Fractal Radiation Pattern Synthesis Technique for Reconfigurable Multiband Arrays," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 7 (July 2003): 1486–1498.
31. Y. Kim and D. L. Jaggard, "The Fractal Random Array," *Proc. IEEE*, vol. 74, no. 9 (1986): 1278–1280.
32. D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning* (Reading, MA: Addison-Wiley Publishing Company, Inc., 1989).
33. R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms* (New York: John Wiley & Sons, Inc., 1998).

34. Y. Rahmat-Samii and E. Michielssen (eds.), *Electromagnetic Optimization by Genetic Algorithms* (New York: John Wiley & Sons, Inc., 1999).
35. J. S. Petko and D. H. Werner, "An Autopolyploidy-based Genetic Algorithm for Enhanced Evolution of Linear Polyfractal Arrays," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 3, part I (March 2007): 583–593.
36. J. Horn, N. Nafpliotis, and D. E. Goldberg, "A Niche Pareto Genetic Algorithm for Multiobjective Optimization," *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, vol. 1 (June 1994): 82–87.
37. N. Srinivas and K. Deb, "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms," *Evolutionary Computation*, vol. 2, no. 3 (Fall 1994): 221–248.
38. C. Fonesca and P. Fleming, "An Overview of Evolutionary Algorithms in Multiobjective Optimization," *Evolutionary Computation*, vol. 3, no. 1 (Spring 1995): 1–16.
39. E. Zitzler and L. Thiele, "An Evolutionary Algorithm for Multiobjective Optimization: The Strength Pareto Approach," *TIK-Report* (Zurich, Switzerland: Swiss Federal Institute of Technology, May 1998).
40. E. Zitzler and L. Thiele, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4 (November 1999): 257–271.
41. J. Knowles and D. Corne, "The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Pareto Multiobjective Optimisation," *1999 Congress on Evolutionary Computation* (July 1999): 98–105.
42. D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective Evolutionary Algorithms: Analyzing the State-of-the-Art," *Evolutionary Computation*, vol. 8, no. 2 (2000) 125–127.
43. M. Laumanns, G. Rudolph, and H. Schwefel, "A Spatial Predator-Prey Approach to Multi-Objective Optimization: A Preliminary Study," *Parallel Problem Solving from Nature—PPSN V*, A. E. Eiben, M. Schoenauer, and H. P. Schwefel (eds). (Amsterdam, Holland: Springer-Verlag, 1998): 241–249.
44. B. Grünbaum and G. C. Shephard, *Tilings and Patterns* (New York: W. H. Freeman and Company, 1987).
45. D. H. Werner, W. Kuhirun, and P. L. Werner, "Fractile Arrays: A New Class of Tiled Arrays with Fractal Boundaries," *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 8 (August 2004): 2008–2018.
46. D. H. Werner, W. Kuhirun, and P. L. Werner, "The Peano-Gosper Fractal Array," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 8 (August 2003): 2063–2072.
47. J. N. Bogard, D. H. Werner, and P. L. Werner, "A Comparison of the Peano-Gosper Fractile Array with the Regular Hexagonal Array," *Microwave and Optical Technology Letters*, vol. 3, no. 6 (December 2004): 524–526.
48. V. Pierro, V. Galdi, G. Castaldi, I. M. Pinto, and L. B. Felsen, "Radiation Properties of Planar Antenna Arrays Based on Certain Categories of Aperiodic Tilings," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 2 (February 2005): 635–644.
49. J. N. Bogard and D. H. Werner, "Optimization of Peano-Gosper Fractile Arrays Using Genetic Algorithms to Reduce Grating Lobes During Scanning," *Proc. IEEE International Radar Conference*, (May 2005): 905–909.
50. J. N. Bogard, D. H. Werner, and P. L. Werner, "Optimization of Peano-Gosper Fractile Arrays for Broadband Performance Using Genetic Algorithms to Eliminate Grating Lobes During Scanning," *Proc. IEEE AP-S Int. Symp.*, vol. 1B (July 2005): 755–758.
51. T. G. Spence and D. H. Werner, "Thinning of Aperiodic Antenna Arrays for Low Side-Lobe Levels and Broadband Operation Using Genetic Algorithms," *Proc. 2006 IEEE AP-S Int. Symp.*, vol. 3 (July 2006): 2059–2062.

Chapter 34

**Low Profile Antenna
Performance
Enhancement
Utilizing Engineered
Electromagnetic
Materials**

Fan Yang

University of Mississippi

Hossein Mosallaei

Northeastern University

Yahya Rahmat-Samii

University of California, Los Angeles

CONTENTS

34.1	INTRODUCTION.....	34-2
34.2	CHARACTERIZATIONS AND DESIGNS OF ELECTROMAGNETIC BAND GAP (EBG) STRUCTURES.....	34-3
34.3	LOW PROFILE WIRE ANTENNAS ON EBG GROUND PLANE.....	34-7
34.4	PATCH ANTENNAS WITH ENHANCED PERFORMANCE USING EBG STRUCTURES.....	34-11
34.5	SURFACE WAVE ANTENNAS WITH A MONOPOLE-LIKE RADIATION PATTERN.....	34-12
34.6	IMPEDANCE AND MAGNETO-DIELECTRIC SUBSTRATES FOR SMALL ANTENNA DESIGNS.....	34-15

34.1 INTRODUCTION

The growing number of wireless communication systems continuously challenges antenna engineers to create novel antenna structures and improve existing antenna designs. Due to the advancements in computational electromagnetics and fabrication technologies, antenna engineers nowadays are capable of exploiting complex engineered electromagnetic materials in antenna designs.¹⁻⁴

Surveying the literature, one observes that various engineered materials have been investigated and numerous antenna applications have been proposed. For example, Figure 34-1 illustrates some representative engineered electromagnetic materials with unique electromagnetic properties.⁵ Frequency selective surfaces (FSS) are widely used in radome and reflector antenna designs as wave filters.^{6,7} Double negative (DNG) material refers to those materials with effective negative permittivity and permeability, which results in properties such as left-handed (LH) wave propagation and negative index of refraction (NIR).⁸⁻¹⁰ Periodic structures that prohibit the propagation of electromagnetic waves in a certain frequency band for certain arrival angles and polarization senses are classified as electromagnetic band gap (EBG) structures.¹¹⁻¹³ Another important category consists of ground planes that exhibit unique reflection characteristics other than conventional PEC, known as complex artificial ground planes.^{14,15}

This chapter summarizes several typical engineered electromagnetic materials and illustrates their applications in antenna engineering. It is demonstrated that they not only improve the performance of conventional antennas such as gain, bandwidth, and efficiency, but also lead to novel radiator concepts and structures like surface wave antennas and reconfigurable antennas.

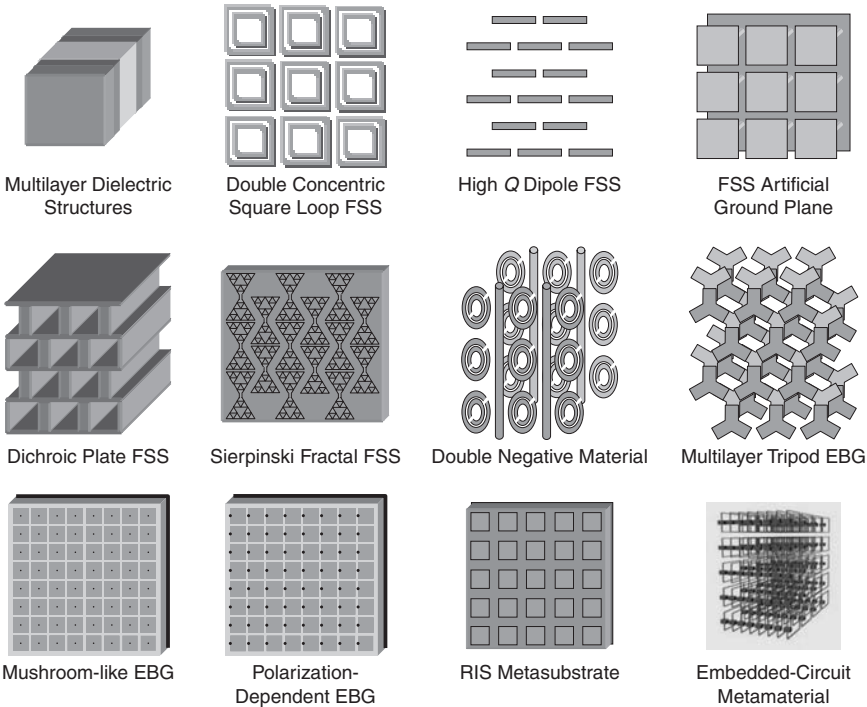


FIGURE 34-1 Different classes of engineered electromagnetic materials

34.2 CHARACTERIZATIONS AND DESIGNS OF ELECTROMAGNETIC BAND GAP (EBG) STRUCTURES

Among various engineered electromagnetic materials, mushroom-like EBG structure has attracted increasing attention.¹² It consists of four parts: a ground plane, a dielectric substrate, periodic metal patches, and connecting vias, as shown in Figure 34-2. This structure is easy to fabricate using the printed circuit board (PCB) technique.

Frequency Band Gap and In-phase Reflection Coefficient

The mushroom-like EBG structure exhibits distinctive electromagnetic properties with respect to incident electromagnetic waves:

- When the incident wave is a surface wave ($k_x^2 + k_y^2 \geq k_0^2$), the EBG structure shows a *frequency band gap* through which the surface wave cannot propagate for any incident angles and any polarization states, resulting in an ideal isolator for electromagnetic waves.
- When the incident wave is a plane wave ($k_x^2 + k_y^2 < k_0^2$), the EBG structure has an *in-phase reflection coefficient* of +1 at a certain frequency, which resembles an ideal perfect magnetic conductor (PMC) that does not exist in nature.

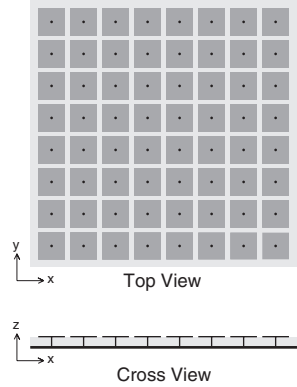


FIGURE 34-2 Geometry of a mushroom-like electromagnetic band gap (EBG) structure (after Yang et al.⁶² © *Microwave Optical and Technology Letters* 2005)

In the above equations, k_x and k_y denote the wave numbers in the horizontal directions, while k_0 is the free-space wave number. The operation mechanism of this EBG structure can be explained by an *LC* filter array: the inductor L results from the current flowing through the vias, and the capacitor C is due to the gap effect between the adjacent patches. Some empirical formulas for the inductor L and the capacitor C are presented in Rahman and Stuchly.¹⁶

To accurately characterize the frequency band gap property and the in-phase reflection coefficient, the finite difference time domain (FDTD) method^{17–20} is used to analyze the EBG structure. A single unit of the EBG structure is simulated with periodic boundary conditions (PBC) incorporated on four sides to model an infinite periodic structure. The dimensions of the analyzed EBG structure are

$$W = 0.10\lambda, g = 0.02\lambda, h = 0.04\lambda, \epsilon_r = 2.94 \quad (34-1)$$

where W is the width of the square patch, g is the gap width, h is the substrate thickness, and ϵ_r is the dielectric constant of the substrate. The vias' radius in the EBG structure is 0.005λ . The free-space wavelength at 4 GHz, $\lambda = 75$ mm, is used as a reference length to define the physical dimensions of the EBG structure. It is worthwhile to point out that the periodicity of the EBG structure (0.12λ) is much smaller than the wavelength. In addition, these dimensions are readily scaled to other frequencies of interest for different applications.

Figure 34-3a shows the ω - β diagram of the EBG structure, where the vertical axis is the frequency and the horizontal axis represents the values of the horizontal wave numbers (k_x, k_y) in the Brillion zone. Each point in the dispersion diagram represents a certain surface wave mode. It is observed that in the frequency range from 3.5 GHz to 5.9 GHz,

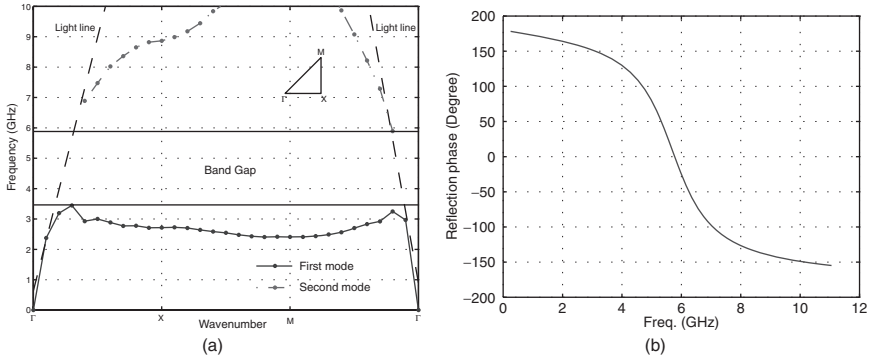


FIGURE 34-3 Electromagnetic properties of an EBG structure characterized using FDTD method with periodic boundary conditions (PBC): (a) the ω - β diagram and (after Yang et al.⁶² *Microwave Optical and Technology Letters* 2005) (b) the reflection phase. The EBG structure exhibits a frequency band gap for surface waves and an in-phase reflection coefficient for plane-wave incidence.

no surface waves can exist regardless of propagation angles or polarization states. Thus, this frequency region is defined as a surface wave band gap of the EBG structure.

Figure 34-3b shows the reflection phase curve for a normally incident plane wave. The reflection phase of a surface is defined as the phase of the reflected E field normalized to the phase of the incident E field at the reflecting surface. It is known that a perfect electric conductor (PEC) has an 180° reflection phase and a perfect magnetic conductor (PMC) has a 0° reflection phase. In contrast, the reflection phase of the EBG surface decreases continuously from 180° to -180° as frequency increases. For example, the EBG surface exhibits a 90° reflection phase around 4.6 GHz and a 0° reflection phase around 5.8 GHz. It is important to note that the reflection phase varies with incident angles and polarization states.

Parametric Studies

The electromagnetic properties of the EBG structure are determined by its physical dimensions, namely, patch width W , gap width g , substrate permittivity ϵ_r , and substrate thickness h . Hence, in-depth studies on these parameters are required to develop engineering design guidelines for the EBG structure.²¹

The reference EBG parameters are listed below:

$$W = 0.12\lambda_{12 \text{ GHz}}, g = 0.02\lambda_{12 \text{ GHz}}, h = 0.04\lambda_{12 \text{ GHz}}, \epsilon_r = 2.20 \quad (34-2)$$

The vias' radius is $0.005\lambda_{12 \text{ GHz}}$. Figure 34-4 presents the effects of the preceding four parameters on the reflection phase of the EBG structure. During each study, only the parameter of interest changes, while the other three parameters remain the same as Eq. 34-2. Two important properties on the reflection phase curve are concerned: the resonant frequency where the reflection phase equals to zero and the slope of the curve that corresponds to the frequency bandwidth of the EBG structure. Following are observations that can be obtained from Figure 34-4:

- When the patch width W increases, the resonant frequency decreases and the slope of the curve becomes steep.
- When the gap width g increases, the resonant frequency increases and the slope of the curve becomes flat.

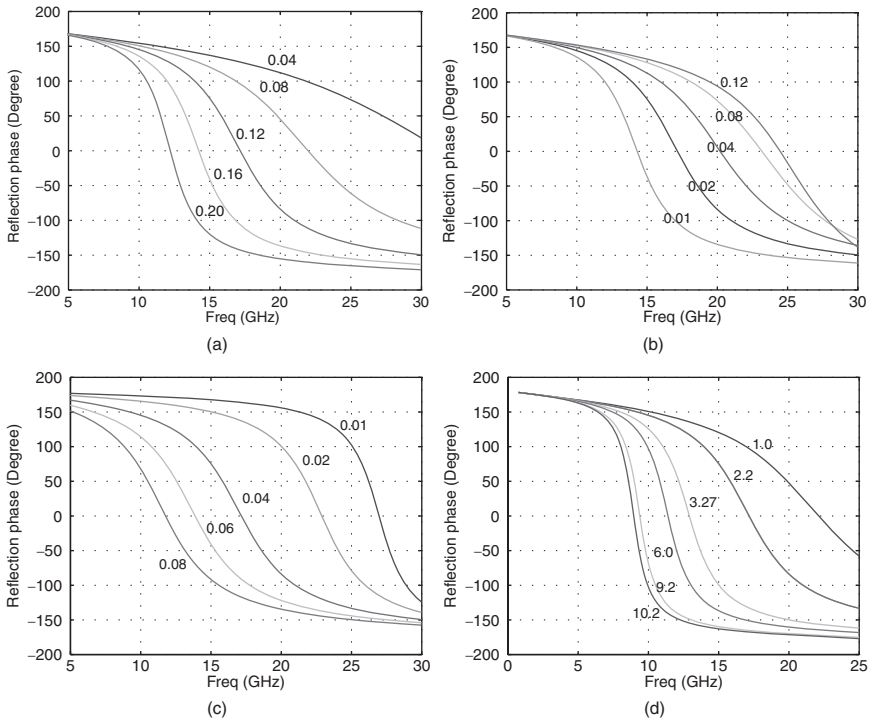


FIGURE 34-4 Parametric studies on the EBG dimensions: (a) patch width W effect, (b) gap width g effect, (c) substrate thickness h effect, and (d) substrate permittivity ϵ_r effect

- When the substrate thickness h increases, the resonant frequency decreases but the slope of the curve becomes flat.
- When the dielectric constant ϵ_r increases, the resonant frequency decreases and the slope of the curve becomes steep.

Although the preceding observations focus on the reflection phase feature of the EBG structure, it is noticed that similar parameter effects also apply in the surface wave band gap property.

Polarization-Dependent EBG Designs

The aforementioned mushroom-like EBG structure has a symmetric square unit cell so that its reflection phase for the normal incidence is independent of the polarization states. When the unit geometry is modified, EBG structures with polarization-dependent reflection phases can be obtained.^{22–24}

Figure 34-5a shows a polarization-dependent EBG design using rectangular patch units. The patch length L is $0.24\lambda_{3\text{ GHz}}$ and the width W is $0.16\lambda_{3\text{ GHz}}$. The gap width is $0.02\lambda_{3\text{ GHz}}$ and the vias' radius is $0.0025\lambda_{3\text{ GHz}}$. The substrate thickness is $0.04\lambda_{3\text{ GHz}}$ and the dielectric constant is 2.20. As a reference, a square patch EBG surface ($0.16\lambda_{3\text{ GHz}} \times 0.16\lambda_{3\text{ GHz}}$) is also studied. Due to different values of L and W , the reflection phase of the EBG surface

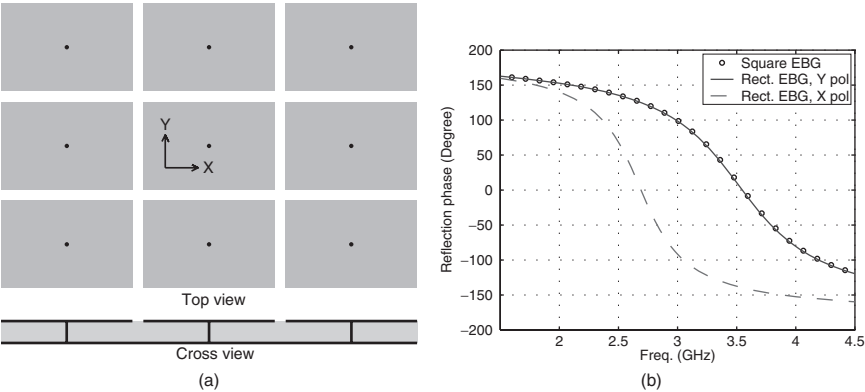


FIGURE 34-5 (a) A rectangular-patch EBG surface and (b) the reflection phases of the EBG structure with respect to different polarizations of the incident wave (after Yang and Rahmat-Samii²³ © *Microwave Opt. Technol. Lett.* 2004)

becomes dependent on the x - or y -polarization state of the incident plane wave, as depicted in Figure 34-5b. When the incident plane wave is y -polarized, the rectangular patch EBG surface has the same reflection phase as the square patch EBG surface because the patch widths are the same. For the x -polarized incident plane wave, the patch length L plays a dominant role in determining the reflection phase. Since the length L is longer than the width W , the reflection phase curve shifts down to lower frequencies. It is noticed that near 3 GHz, the EBG surface shows a -90° reflection phase for the x -polarized wave and a $+90^\circ$ reflection phase for the y -polarized wave.

Besides the rectangular patch unit, various other approaches have also been used to realize the polarization-dependent feature. For example, Figure 34-6 shows two alternative designs:

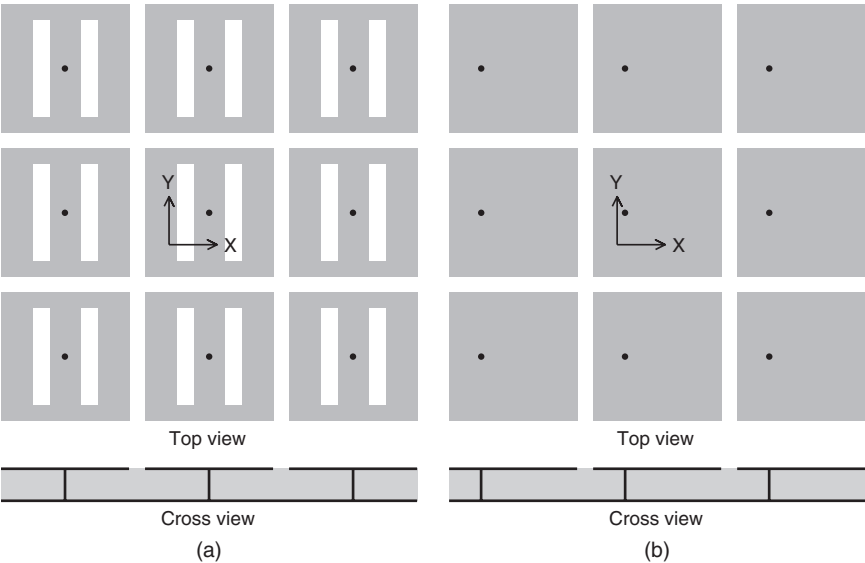


FIGURE 34-6 Polarization-dependent EBG structures: (a) slot-loaded design and (b) offset vias design (after Yang and Rahmat-Samii²³ © *Microwave Opt. Technol. Lett.* 2004)

a slot-loaded EBG and an EBG with offset vias. In Figure 34-6a, a pair of y-oriented slots are symmetrically incorporated into the patch of a square EBG surface. The slots affect electric currents flowing along the x direction, resulting in a longer current path. Thus, the reflection phase of the x -polarized wave decreases to a lower frequency. In contrast, the reflection phase of the y -polarized wave remains the same as the square patch EBG surface because the slots have little effect on the electric currents flowing along the y direction.

In Figure 34-6b the vias are offset along the x direction while they are still centered along the y direction. Therefore, the reflection phase for the y -polarized wave remains unchanged whereas the reflection phase for the x -polarized wave varies with the vias' position. When the vias are located in the center of the patch, only one resonant frequency is observed. When the vias are offset, dual resonance behaviors are observed with one resonant frequency higher than the original frequency and the other lower. The different frequencies correspond to the different widths of the left and right regions of the patch with respect to the vias. The left part is narrower and thus corresponds to the higher resonant frequency. The right part is wider and is related to the lower resonant frequency.

34.3 LOW PROFILE WIRE ANTENNAS ON EBG GROUND PLANE

Because of the attractive electromagnetic properties, the EBG structure has been applied in various antenna applications. In this section, the in-phase reflection coefficient feature is exploited to increase the radiation efficiency of low profile wire antennas.^{21,25-27} The low profile design usually refers to an antenna structure whose overall height is less than one-tenth of the operating wavelength, which is desirable in many mobile communication systems.

Dipole Antenna on an EBG Ground Plane

To appreciate the advantages of using an EBG ground plane, we start with a dipole antenna, the simplest radiating structure. As shown in Figure 34-7a, a dipole antenna is horizontally positioned above an EBG ground plane to obtain a low profile configuration.²⁸ The dipole length is $0.40\lambda_{12\text{ GHz}}$ and its radius is $0.005\lambda_{12\text{ GHz}}$. The height of the dipole over the top surface of the EBG ground plane is only $0.02\lambda_{12\text{ GHz}}$. A finite EBG ground plane with a $1\lambda_{12\text{ GHz}} \times 1\lambda_{12\text{ GHz}}$ size is used in the analysis, and the dimensions of the EBG structure are the same as those given in Eq. 34-2.

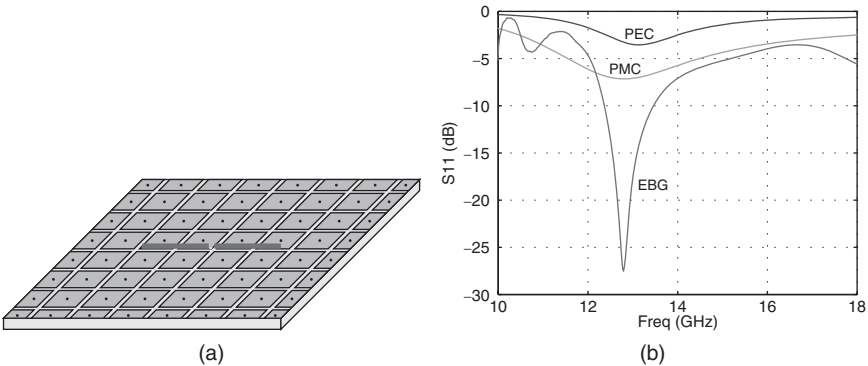


FIGURE 34-7 Dipole antenna near an EBG ground plane: (a) geometry and (b) FDTD simulated return loss (after Yang and Rahmat-Samii²¹ © IEEE 2003)

Figure 34-7*b* shows the FDTD simulated return loss of the dipole antenna. For comparison purposes, the performances of dipole antennas on PEC and PMC ground planes are also simulated and plotted. When the dipole is located above the PEC ground plane, the return loss is only -3.5 dB. The PEC surface has a 180° reflection phase so that the direction of the image current is opposite to that of the original dipole. The radiations from the image current and the original dipole cancel each other, resulting in a very poor return loss. When the PMC surface, which has a reflection phase of 0° , is used as the ground plane, the dipole has a return loss of -7.2 dB. The return loss is improved; however, it suffers from a strong mutual coupling between the image current and the dipole due to their close proximity.^{21,29} The best return loss of -27 dB is achieved by the dipole antenna over the EBG ground plane. The reflection phase of the EBG varies with frequency from 180° to -180° . In a certain frequency range, the EBG surface successfully serves as the ground plane for a low profile dipole, resulting in a high radiation efficiency. It is clear from this comparison that the EBG surface is a good ground plane candidate for low profile wire antenna designs.

A further question is how to identify the operational frequency band of an EBG ground plane. To this end, the parameters of the EBG surface are fixed and the length of the dipole is varied to resonate at different frequencies. By observing the return loss value and radiation patterns of the dipole at different frequencies, one can find a useful operational frequency band of the EBG ground plane for low profile wire antenna designs. Figure 34-8*a* shows the return loss results of a dipole with its length varying from $0.26 \lambda_{12 \text{ GHz}}$ to $0.60 \lambda_{12 \text{ GHz}}$. It is observed that the dipole shows a return loss better than -10 dB in a frequency range from 11.5 to 16.6 GHz.

From a computational efficiency viewpoint, it would be interesting to know if one could directly use the reflection phase curve of the EBG structure to identify the operational frequency band. Thus, the reflection phase of the EBG surface is evaluated, and the result is presented in Figure 34-8*b*. In contrast to the 180° reflection phase of a PEC surface or the 0° reflection phase of a PMC surface, if one chooses the $90^\circ \pm 45^\circ$ reflection phases as the criterion for the EBG ground plane, a frequency region from 11.3 to 16 GHz is identified, which is close to the frequency region obtained in the dipole model. It is revealed from this comparison that the operational frequency band of an EBG ground plane is the frequency region inside which the EBG surface shows a quadratic reflection phase ($90^\circ \pm 45^\circ$). This quadratic reflection phase criterion has been further demonstrated by various numerical and experimental results.²¹

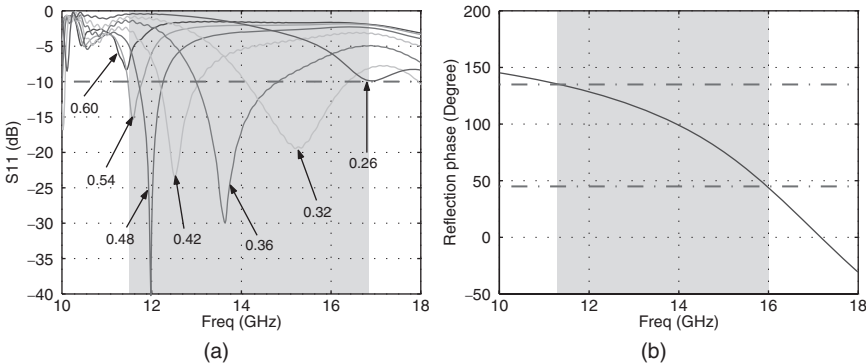


FIGURE 34-8 Identification of the operational frequency band of an EBG ground plane: (a) return loss of a nearby dipole with length varying from 0.26 to $0.60 \lambda_{12 \text{ GHz}}$ and (b) reflection phase of the EBG surface (after Yang and Rahmat-Samii²¹ © IEEE 2003)

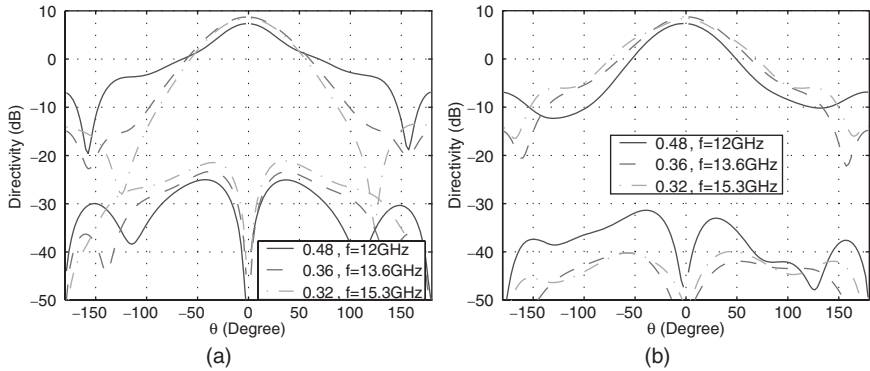


FIGURE 34-9 Radiation patterns of three dipoles at their resonant frequencies: (a) *E*-plane patterns and (b) *H*-plane patterns (after Yang and Rahmat-Samii²¹ © IEEE 2003)

The radiation patterns of dipole antennas above the EBG ground plane are also calculated to verify the radiation efficiency. Figure 34-9 displays both the *E*- and *H*-plane patterns of three dipole antennas at their resonant frequencies: (1) $0.48 \lambda_{12 \text{ GHz}}$ dipole resonating at 12 GHz, (2) $0.36 \lambda_{13.6 \text{ GHz}}$ dipole resonating at 13.6 GHz, and (3) $0.32 \lambda_{15.3 \text{ GHz}}$ dipole resonating at 15.3 GHz. It is observed that all three dipoles radiate efficiently with directivities around 8 dB.

Circularly Polarized Designs

Circularly polarized (CP) antennas are desired in many communication systems such as the Global Positioning System (GPS) and satellite links. The EBG ground plane is also implemented in low profile CP wire antenna designs. The first approach is to replace the traditional PEC ground plane of a curl antenna with an EBG ground plane, as shown in Figure 34-10a. The circular polarization is generated by the traveling current along the curl and the EBG ground plane helps to improve the radiation efficiency of the curl in a low profile configuration.^{31,32} It is noticed that the overall antenna height (0.07λ)

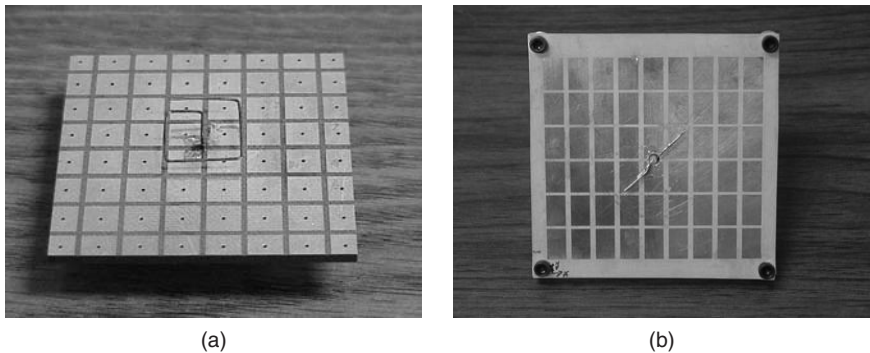


FIGURE 34-10 Photographs of low profile CP wire antennas on EBG ground plane: (a) a curl on a square patch EBG ground plane (after Yang and Rahmat-Samii³¹ © *Microwave Opt. Technol. Lett.* 2001) and (b) a dipole on a rectangular patch EBG ground plane (after Yang and Rahmat-Samii³³ © IEEE 2005)

is much smaller than that of a conventional curl design on a PEC ground plane (0.25λ). Figure 34-10b presents the second approach for low profile CP wire antenna design where a linearly polarized dipole is used instead of the CP curl.³³ The circular polarization pattern of the overall antenna is realized by the superposition of the directly radiating wave from the dipole and the reflected wave from the EBG ground plane. A polarization-dependent EBG ground plane is carefully designed and implemented here so that the reflected wave has a perpendicular polarization direction to the directly radiating wave as well as a 90° phase shift.²³ The function of the polarization-dependent EBG ground plane is similar to a meandering line polarizer.⁶

Reconfigurable Antenna with Pattern Diversity

Another interesting application of the EBG ground plane is to build a low profile reconfigurable wire antenna with radiation pattern diversity.³⁴ Reconfigurable antennas are desirable in modern wireless communication systems^{35–37} because they can provide more functionalities than ordinary antennas by reconfiguring their radiation performance, such as the operating frequencies,^{38,39} polarizations,^{40–42} and radiation patterns.^{43–47} Figure 34-11a shows the geometry of a reconfigurable wire antenna design, where a feeding probe is connected to two metal strips through two switches. When the left switch is ON and the right switch is OFF, the probe has an electrical connection to the left strip, resulting in a bent monopole oriented along the $-x$ direction. When the left switch is OFF and the right switch is ON, the probe has an electrical connection to the right strip, resulting in a bent monopole oriented along the $+x$ direction. As a consequence, the direction of the antenna beam can be switched in space and the diversity in the radiation pattern is realized. A reconfigurable antenna prototype is built to demonstrate the operational mechanism, and the measured diversity patterns are shown in Figure 34-11b. It is observed that the antenna beam is switched between $\pm 26^\circ$.

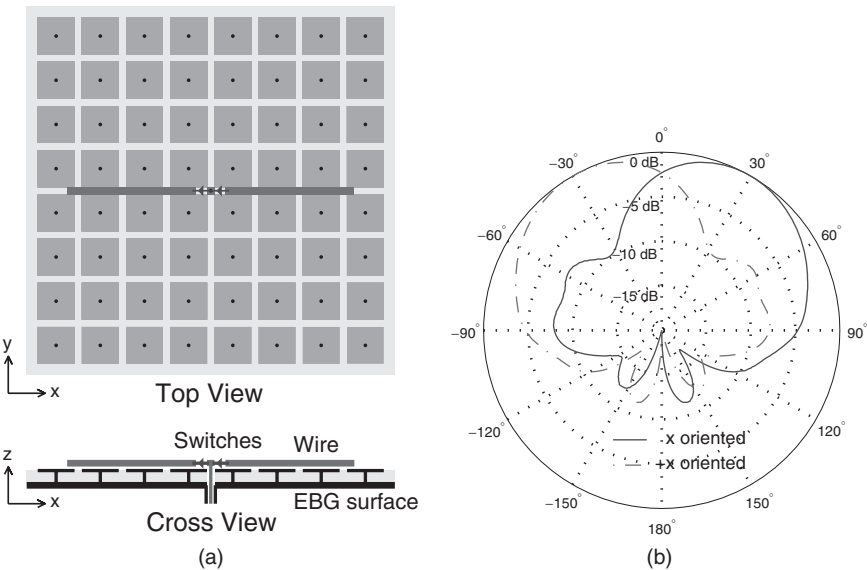


FIGURE 34-11 A low profile reconfigurable wire antenna on the EBG ground plane with radiation pattern diversity: (a) antenna geometry and (b) measured radiation patterns

34.4 PATCH ANTENNAS WITH ENHANCED PERFORMANCE USING EBG STRUCTURES

The EBG structure has two main features, and the in-phase reflection coefficient feature is used in the preceding section to improve the radiation efficiency of low profile wire antennas. This section discusses how the surface wave band gap feature is used to enhance the performance of microstrip antennas and arrays. Microstrip antennas are widely used in wireless communications due to their numerous advantages, including low profile configuration, light weight, low fabrication cost, and conformability with RF circuitry. In typical microstrip antenna designs, unwanted surface waves are excited in the substrate. These surface waves degrade the antenna performance, such as decreasing the antenna gain, increasing the back lobe, and increasing the mutual coupling. This problem becomes more severe when high dielectric constant substrates are used to design miniaturized microstrip antennas. Various methods have been proposed to solve this problem,^{48–51} and the EBG structures demonstrate excellent capability in reducing the surface waves.^{52–61} In this section, the mushroom-like EBG structure is integrated into microstrip antennas and arrays, and the applications of metasubstrate in microstrip antennas will be discussed in Section 34.6.

Patch Antenna Element

The EBG structure is first integrated with a microstrip patch antenna element to increase the antenna gain and suppress the back radiation.⁵⁴ The basic idea is to properly design the EBG structure such that the resonant frequency of the patch antenna falls inside the band gap of the EBG structure, hence inhibiting the surface waves. Figure 34-12a shows a patch antenna surrounded by four rows of EBG cells. The patch antenna is designed to resonant at 5.8 GHz and is fabricated on a 2.54-mm-thick RT/duroid 6010 ($\epsilon_r = 10.2$) substrate with a finite ground plane of 52×52 mm ($1\lambda \times 1\lambda$ at 5.8 GHz). The EBG patch size is 2.5×2.5 mm and the gap width between adjacent patches is 0.5 mm. It is noticed that the EBG structure is compact in this design because a thick substrate with a high dielectric constant is used. For comparison purposes, a conventional patch antenna on the same substrate but without EBG cells is built as a reference. It is noticed that both antennas have similar return loss,

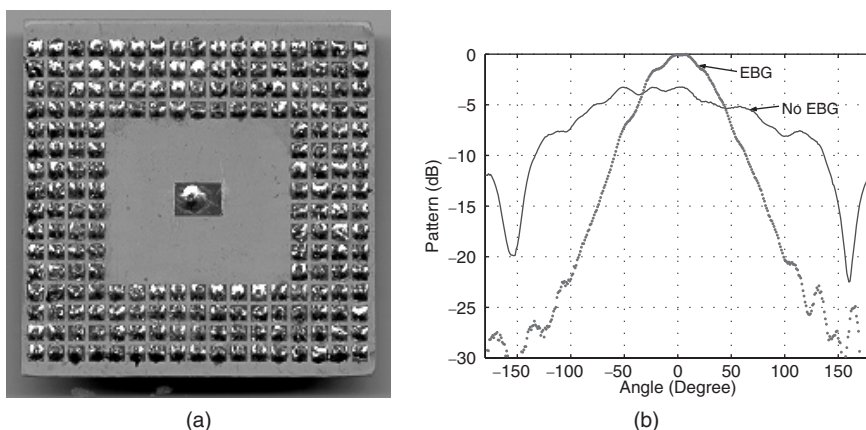


FIGURE 34-12 (a) Photo of a microstrip patch antenna surrounded by an EBG structure and (b) measured *E*-plane radiation patterns for patch antennas with and without EBG structure

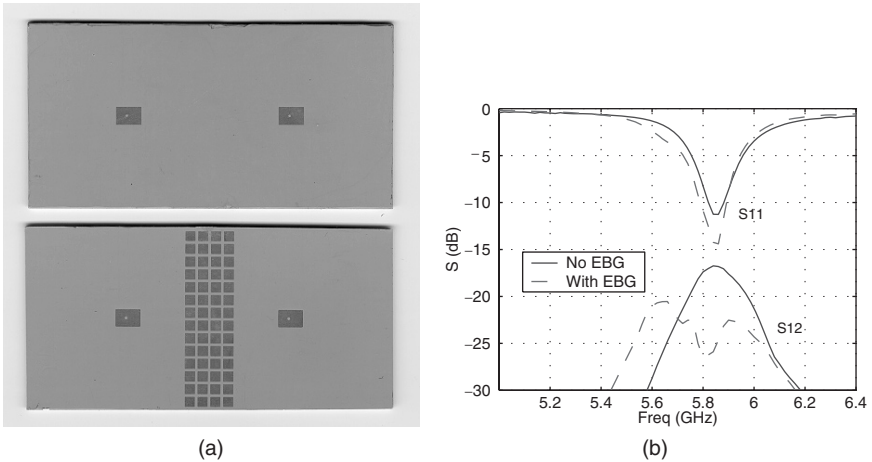


FIGURE 34-13 (a) Photos of a normal microstrip antenna array (above) and a design that integrates the EBG structure into the array (below); (b) measured scattering coefficients of the microstrip antenna arrays (after Yang and Rahmat-Samii⁵⁵ © IEEE 2003)

but the EBG case shows an improved radiation pattern, as plotted in Figure 34-12b. With the surrounding EBG cells, the antenna gain is successfully increased by 3.2 dB while the back lobe is effectively reduced by more than 15 dB.

Patch Antenna Array

The EBG structure also can be integrated with a microstrip antenna array to reduce the mutual coupling between antenna elements.⁵⁵ As shown in Figure 34-13a, two pairs of microstrip antennas with and without the EBG structure are fabricated on 1.92-mm-thick Roger RT/duroid 6010 substrates. The antenna's size is 6.8×5 mm, and they are fabricated on a finite ground plane of 100×50 mm with the patch distance of 38.5 mm ($0.75 \lambda_{5.8 \text{ GHz}}$). The EBG patch width is 3 mm and the gap width is 0.5 mm.

The measured results are shown in Figure 34-13b, where both antennas resonate at 5.86 GHz with a return loss better than -10 dB. For the antenna array without the EBG structure, the mutual coupling at 5.86 GHz is as high as -16.8 dB. In contrast, the mutual coupling of the antennas with the EBG structure is only -24.6 dB. An 8-dB reduction of mutual coupling is achieved at the resonant frequency because of the effective suppression of surface waves by the EBG structure. This low mutual coupling design is potentially useful for various phase array applications, such as eliminating blind angles in radar systems.

34.5 SURFACE WAVE ANTENNAS WITH A MONOPOLE-LIKE RADIATION PATTERN

The mushroom-like EBG structure exhibits both a surface wave band gap feature and an in-phase reflection coefficient. However, these two characteristics are not necessarily associated with each other.⁶² For example, this section discusses another engineered electromagnetic material that only has an in-phase reflection coefficient but has no band gap for surface waves. This new engineered material is used to design surface wave antennas that realize a monopole-like radiation pattern with a low profile configuration.

Dipole-Fed Surface Wave Antenna

Figure 34-14a shows the geometry of a horizontal dipole antenna above a patch-loaded grounded slab. The dipole is simply fed by a 50- Ω coaxial cable: one arm of the dipole is connected to the center probe of the cable, and the other arm is connected to the outside conductor of the cable, which is soldered to the lower PEC of the artificial surface. Compared to Figure 34-7a, this design uses a patch-loaded grounded slab as the ground plane, which is similar to a mushroom-like EBG structure except that the vertical vias are all removed. It is revealed that removing the vertical vias has little effect on the in-phase reflection feature for a normally incident plane wave; however, it completely changes the surface wave characteristic of the engineered ground plane. Figure 34-14b presents the dispersion diagram of the patch-loaded grounded slab. The dimensions of this artificial ground plane are the same as those given in Eq. 34-1. It is clear that when the vertical vias are removed, the surface wave band gap disappears. Therefore, the surface waves can exist over the entire frequency band.

Due to the different properties of the engineered ground plane, the radiation performance of the antenna is vastly different from the ones discussed in Section 34.3. The return loss of the antenna with various dipole lengths is shown in Figure 34-15a. It is noticed that the low profile antenna also obtains a good return loss in a frequency region where the artificial surface exhibits a quadratic reflection phase. When the dipole length is $0.26\lambda_{4\text{ GHz}}$, the antenna achieves a good return loss of -30 dB at 4.05 GHz with a 7.1 percent impedance bandwidth. It is worthwhile to point out that the length of the dipole is much smaller than the half wavelength at the operating frequency. In contrast, if a dipole is located near an EBG ground plane and resonates at the same frequency, the length of the dipole is close to a half wavelength.

Figure 34-15b shows the radiation patterns of the antenna at the resonant frequency of 4.05 GHz. Several interesting observations can be made from this figure. First, the antenna shows a small radiation power in the broadside direction ($\theta = 0^\circ$). The main beam of this antenna points to the $\theta = 50^\circ$ direction with a directivity of 5 dBi. This is different from the

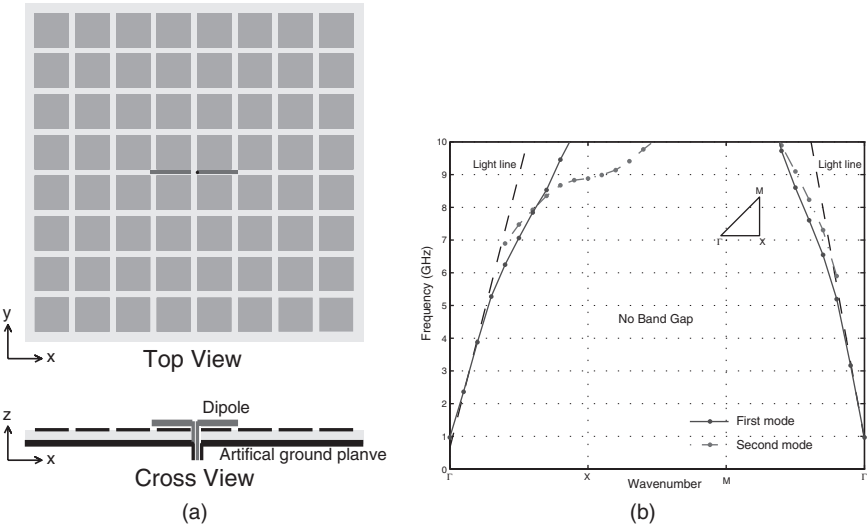


FIGURE 34-14 (a) A dipole antenna near a patch-loaded grounded slab and (b) the dispersion diagram of the patch-loaded grounded slab. No surface wave band gap exists for this engineered structure (after Yang et al.⁶² © *Microwave Optical and Technology Letters* 2005).

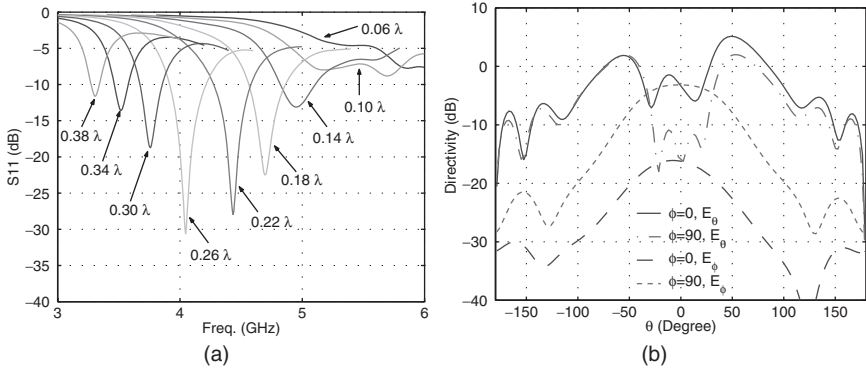


FIGURE 34-15 Radiation characteristics of the dipole-fed surface wave antenna: (a) FDTD simulated return loss of the antenna with different dipole lengths (after Yang et al⁶² © *Microwave Optical and Technology Letters* 2005) and (b) radiation patterns of the antenna at 4.05 GHz with a dipole length of 0.26λ

dipole antenna on an EBG ground plane whose main beam points to the broadside direction, as shown in Figure 34-9. Second, E_θ is the copolarized field in both the xz ($\phi = 0^\circ$) and yz planes ($\phi = 90^\circ$). Thus, both the xz and yz planes are E planes. In contrast, if the dipole is near an EBG ground plane, the xz plane is the E plane but the yz plane is the H plane. In summary, this antenna has entirely different radiation characteristics compared to a horizontal dipole on an EBG ground plane.

The radiation characteristics in Figure 34-15 can be explained by the propagation and diffraction of surface waves in the engineered ground plane. When a dipole is positioned near a patch-loaded grounded slab, strong surface waves are excited to propagate along the ground plane. The dipole works more like a transducer than a radiator. Therefore, the optimal length of the dipole is not necessarily equal to a half wavelength. It is also noticed that the surface waves are dominated by the TM_z mode and the electric field is vertically polarized. When the surface waves diffract at the edge of the ground plane, a monopole-like radiation pattern is generated. For example, since the diffraction at the edge is hard boundary diffraction, two diffracted rays from opposite edges will cancel in the broadside direction, resulting in a radiation null. The hard diffraction also determines the polarization of the radiating field along the θ direction. Therefore, this antenna is properly identified as a surface wave antenna (SWA)^{63–67} because of this operational mechanism.

The attractive feature of this SWA design is the low profile configuration. The height of the horizontal dipole over the artificial ground plane is only $0.02\lambda_{4\text{ GHz}}$ whereas the height of a typical monopole antenna is $0.22\lambda_{4\text{ GHz}}$. The dipole height is less than 10 percent of the conventional monopole antenna. Therefore, this low profile SWA design has a promising potential in mobile communication systems such as vehicle radio systems.

Patch-Fed Surface Wave Antenna

A deficiency in the previous design is the relatively high cross-polarization attributed to the direct radiating field from the dipole. To solve this problem and obtain a better monopole-like radiation pattern, a patch-fed surface wave antenna (PFSWA) is sketched in Figure 34-16a. The same patch-loaded grounded slab is used in the antenna design while a circular patch is inserted in the middle of the substrate to excite surface waves.⁶⁸ Investigation shows that the energy can be effectively coupled from the cavity fields underneath the circular patch to the surface waves propagating along the engineered ground plane. Hence, good radiation

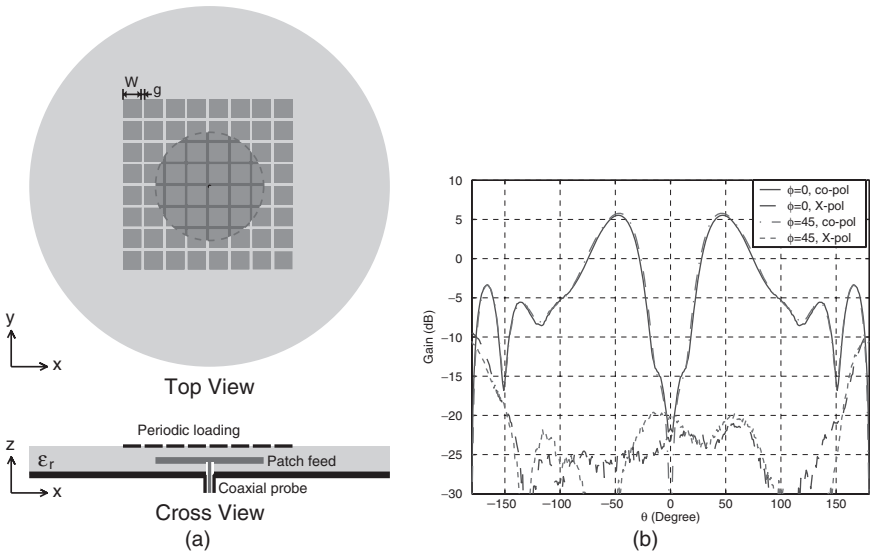


FIGURE 34-16 A low profile patch-fed SWA with a monopole-like radiation pattern: (a) antenna geometry and (b) measured radiation patterns

efficiency can be obtained. Parametric studies reveal that 8×8 periodic patches are enough to launch the surface waves. In order to get a symmetric diffraction pattern, the ground plane is truncated into a circular shape. Figure 34-16b shows measured radiation patterns of a PFSWA prototype, which is very similar to a typical monopole pattern. Both the xz plane ($\phi = 0^\circ$) and diagonal plane ($\phi = 45^\circ$) patterns are presented, which are almost identical to each other. It has a deep null in the broadside direction and the antenna beam is at $\theta = 47^\circ$ direction with a gain of 5.6 dBi. The co-polarization is along the θ direction, and the cross-polarization is -25 dB lower than the co-polarization in the front side. A noticeable cross-polarization is observed in the backside of the antenna due to diffractions from the supporting posts and feeding cable.

34.6 IMPEDANCE AND MAGNETO-DIELECTRIC SUBSTRATES FOR SMALL ANTENNA DESIGNS

The growing number of wireless applications has presented RF engineers with a continuing demand for low-cost, power-efficient, and small-size system designs. Depending on the application at hand and required system characteristics, such as data rate, environment, range, etc., the system parameters, such as operating frequency, transmitter power, and modulation scheme, may vary widely. However, independent of the application, compactness, wide bandwidth, high efficiency, ease of fabrication and integration, and low cost are always sought in wireless systems. One of the most important components of wireless systems is their antenna. The substrate of the antenna has a significant role in successfully providing the above engineering requirements. In this section, three novel substrates—namely, reactive impedance, layered magneto-dielectric, and embedded-circuit metasubstrates—are presented, and their advantages and challenges in the design of small antennas with improved bandwidth performance are addressed.

Reactive Impedance Surface

Design of antenna elements with significant front-to-back radiation ratio is usually accomplished through the use of metal-backed substrates.⁶⁹ However, printed antennas on metal-backed substrates have limited bandwidth and efficiency. As demonstrated earlier, this problem stems from the fact that the radiated field from the image of the antenna's electric current, which is placed in close proximity and parallel to a PEC, tends to cancel out the radiated field from the antenna current itself. In this case, matching the antenna input impedance is rather difficult, and if a matching condition can be achieved, it would be over a relatively narrow bandwidth. To circumvent this difficulty, a reactive impedance surface (RIS), as proposed in Mosallaei and Sarabandi,²⁹ can be used for the antenna substrate. The RIS has three major features:

- It provides a total reflection power that enhances the antenna front-to-back ratio.
- The image of a point source located above the RIS is a spatially distributed current element that has the minimum interaction with the point source. This has the significant advantage of reducing the mutual coupling between the antenna and its substrate, resulting in the impedance bandwidth enhancement. Mutual coupling F between an infinitesimal dipole located at $0.02\lambda_0$ above the RIS and its image as a function of the normalized impedance ν/η_0 is demonstrated in Figure 34-17. It shows the minimum coupling occurs for a surface impedance of $\nu = 0.33\eta_0$, an impedance between the PEC and PMC behaviors, as pointed out before.
- RIS has the ability to store magnetic (or electric) energy that can be properly used to compensate for the near-field electric (or magnetic) energy of the radiating structure resulting in the antenna size reduction.

To realize an RIS with surface impedance $\eta = j\nu$, a periodic configuration of small-size patch elements printed on a dielectric material backed by a PEC can be used (see Figure 34-18a). The gap capacitances between the patches provide an equivalent capacitor that is in parallel with an inductor obtained from the PEC after the distance d through the dielectric material. A parallel LC circuit offers a reactive impedance behavior satisfying the required condition. The FDTD is applied to demonstrate the surface impedance property of the RIS and the results for normal and oblique incident waves are plotted in Figure 34-18b. Since the dielectric constant of substrate is relatively large,

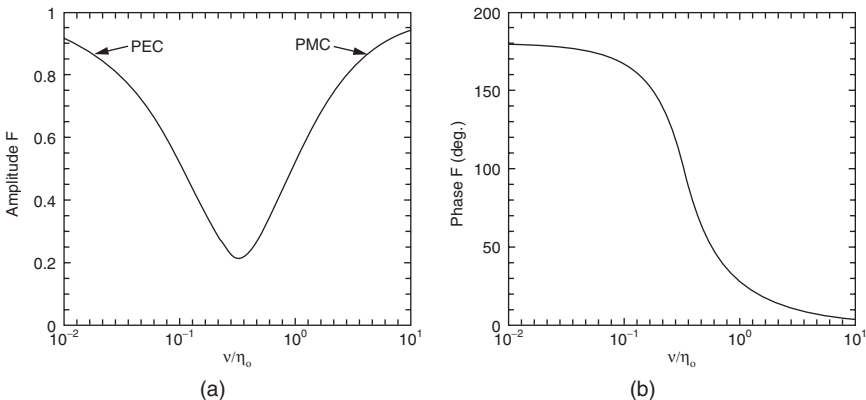


FIGURE 34-17 Mutual coupling between an infinitesimal dipole, located above the RIS, and its image: (a) amplitude and (b) phase (after Mosallaei and Sarabandi²⁹ © IEEE 2004)

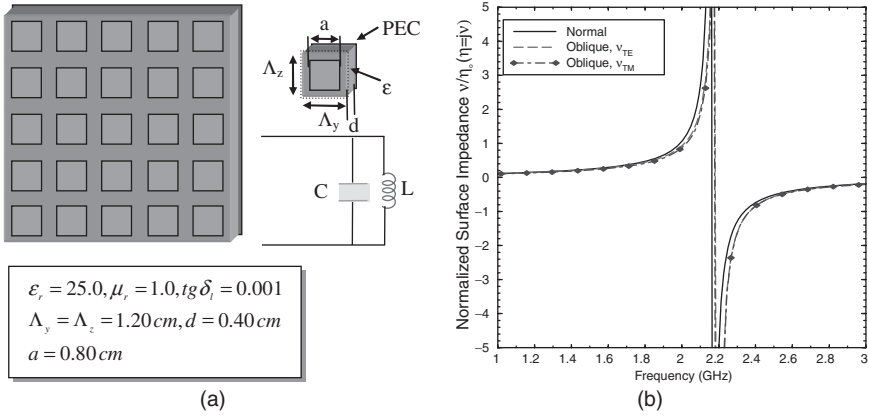


FIGURE 34-18 An RIS constructed from the periodic array of patches printed on a dielectric material backed by PEC: (a) the geometry and its equivalent circuit model and (b) normalized surface impedance (after Mosallaei and Sarabandi²⁹ © IEEE 2004)

according to Snell's law, the transmission angle is almost independent of the incident angle. As a result, the surface performance is almost insensitive to the incident angle (see Figure 34-18b).

It must be mentioned that in the proposed design, no connecting vias between the patch elements and ground plane are constructed. In fact, the vias have nothing to do with the reflection phase and impedance performance of the structure. The major application of vias is when undesirable surface waves can propagate through the substrate, where in this case the vias can be used to successfully forbid the propagation of these waves. This has been clearly demonstrated in Section 34.4.

Figure 34-19a shows the geometry of a patch antenna printed on a relatively low dielectric material with $\epsilon_r = 6$ and located above the RIS substrate. The FDTD is applied to characterize the structure, and the return loss is determined in Figure 34-19b, where a resonant frequency of $f_0 = 1.86 \text{ GHz}$ is obtained. The inductive property of RIS below the resonant frequency is properly combined with the capacitive behavior of patch in this spectral range and successfully shifts down the resonance frequency of antenna, resulting in the antenna

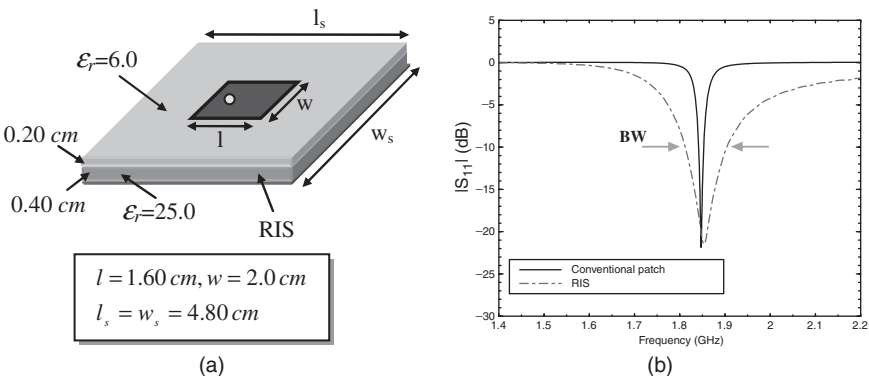


FIGURE 34-19 (a) Patch antenna printed on RIS and (b) its return loss (after Mosallaei and Sarabandi²⁹ © IEEE 2004)

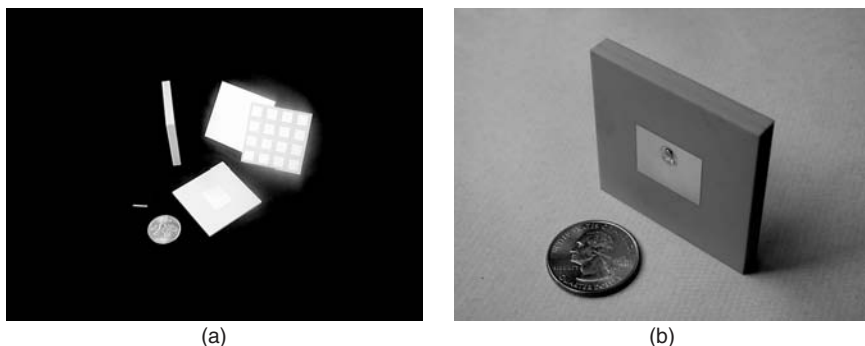


FIGURE 34-20 Fabrication of patch on the RIS substrate: (a) magnesium silicate and magnesium calcium titanate blocks metalized and etched and (b) assembled patch antenna over the RIS substrate (after Mosallaei and Sarabandi²⁹ © IEEE 2004)

miniaturization by a factor of about $5(\lambda_0/10)$ antenna size). In addition, since the RIS has a very small coupling with the antenna, a relatively wide bandwidth of about $BW = 5.0\%$ is achieved. For comparison, the performance of the patch printed on a dielectric material with $\epsilon_r = 21.0$ (providing the same resonant frequency) and thickness $t = 0.2$ cm is also plotted in Figure 34-19b. Patch radiator printed on the conventional substrate provides a very narrow bandwidth of about $BW = 0.63\%$. A thicker conventional dielectric substrate may provide a wider bandwidth; however, the impedance matching is more difficult.

Fabrication of the patch antenna printed over the RIS substrate is illustrated in Figure 34-20. Two independent layers are fabricated separately on high-quality ceramic substrates. The dielectric material used for the patch antenna substrate is Trans-Tech D-6 magnesium silicate ($\epsilon_r = 6$), commonly known as Forsterite. The RIS substrate is made using Trans-Tech MCT-25 magnesium calcium titanate composition ($\epsilon_r = 25$). Using a thick film silver paste and Trans-Tech's screen printing process, the array of square patches, corresponding metal backing, and patch are generated. Both substrates are heat treated to form an intimate bond of the silver to the dielectric material. This intermediary stage is diagrammed in Figure 34-20a. Using a two-part, low-loss dielectric adhesive, the substrates are assembled in a fixture to ensure alignment. The return loss of the fabricated patch antenna is measured and plotted in Figure 34-21a. The antenna resonance is found to be at $f_0 = 1.92$ GHz and it exhibits an impedance match with better than -25 -dB return loss. The measured relative bandwidth is $BW = 6.71\%$. The radiation patterns are measured in the anechoic chamber of the University of Michigan Radiation Laboratory and are shown in Figure 34-21b. The gain and front-to-back ratio are, respectively, measured to be $G = 4.5$ dBi and 5.6 dB. This measured gain corresponds to an excellent radiation efficiency of $\epsilon_r = 90\%$. To our knowledge, this is the highest reported gain and bandwidth for such a small planar antenna.

Magneto-Dielectric Substrate

In this section, the design of a layered magneto-dielectric substrate for antenna miniaturization and bandwidth enhancement is presented. A common approach for antenna miniaturizing is to print the radiator on a high-permittivity substrate. However, the strong electric energy stored inside the high dielectric substrate beneath the patch increases the radiation quality factor of antenna (Q) resulting in a narrow impedance bandwidth. To overcome this problem, one can use a magneto-dielectric material with moderate values of μ_r and ϵ_r .

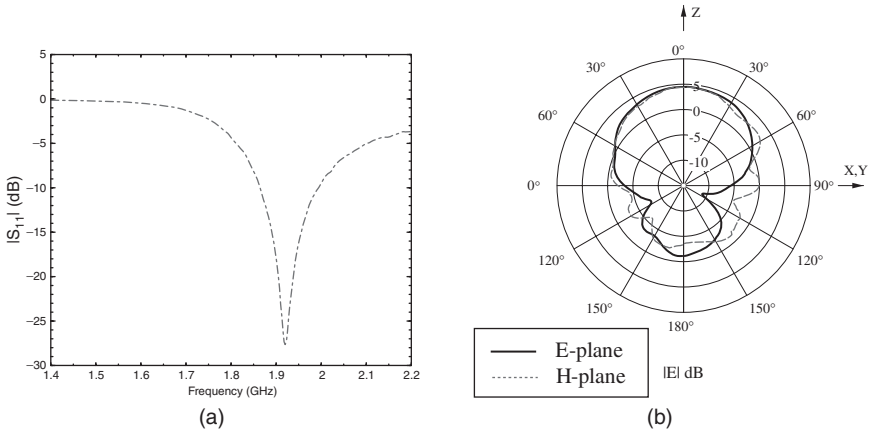


FIGURE 34-21 (a) Measured return loss and (b) radiation patterns of the patch antenna fabricated over the RIS substrate (after Mosallaei and Sarabandi²⁹ © IEEE 2004)

that provides the same miniaturization factor $n = \sqrt{\mu_r \epsilon_r}$, while a wider bandwidth can be achieved.⁷⁰ It has been shown by Hansen and Burke⁷¹ that the zero-order bandwidth for an antenna over a magneto-dielectric substrate with thickness t can be approximated by

$$BW \approx \frac{96\sqrt{\mu_r/\epsilon_r} t/\lambda_0}{\sqrt{2[4 + 17\sqrt{\mu_r \epsilon_r}]}}. \quad (34-3)$$

Thus for a given miniaturization factor (constant $\sqrt{\mu_r \epsilon_r}$), the antenna bandwidth can be enhanced by increasing μ_r/ϵ_r ($\mu_r > \epsilon_r$). Figure 34-22a shows a patch antenna with size 10×8 cm printed on a four-layer dielectric and hexaferrite materials with thickness 2 cm. The size of the ground plane is 20×20 cm. The hexaferrite is a Z-type hexagonal material fabricated in Trans-Tech. It has the permittivity of around 16 and permeability shown in Figure 34-22b. The dielectric material has permittivity $\epsilon_r = 2.2$ and loss tangent $tg\delta_l = 0.001$. Notice that to offer a low-loss radiation performance, the antenna must operate away from the material resonant frequency.

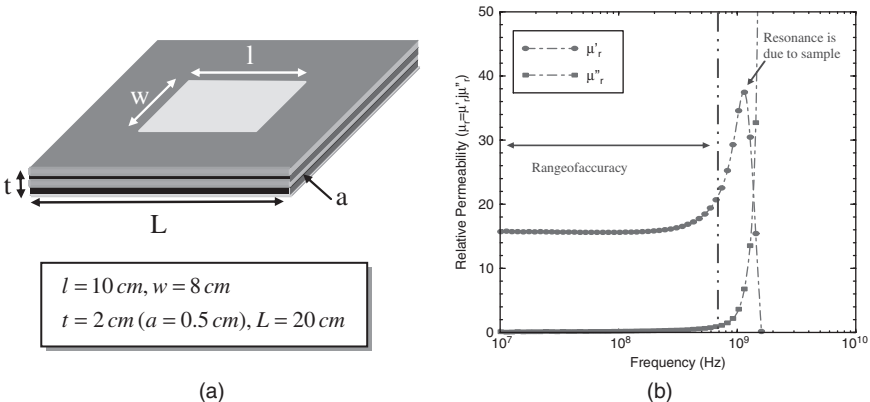


FIGURE 34-22 Patch antenna printed on four-layer dielectric and hexaferrite substrate: (a) the geometry and (b) hexaferrite performance (after Mosallaei and Sarabandi⁷⁰ © IEEE 2004)

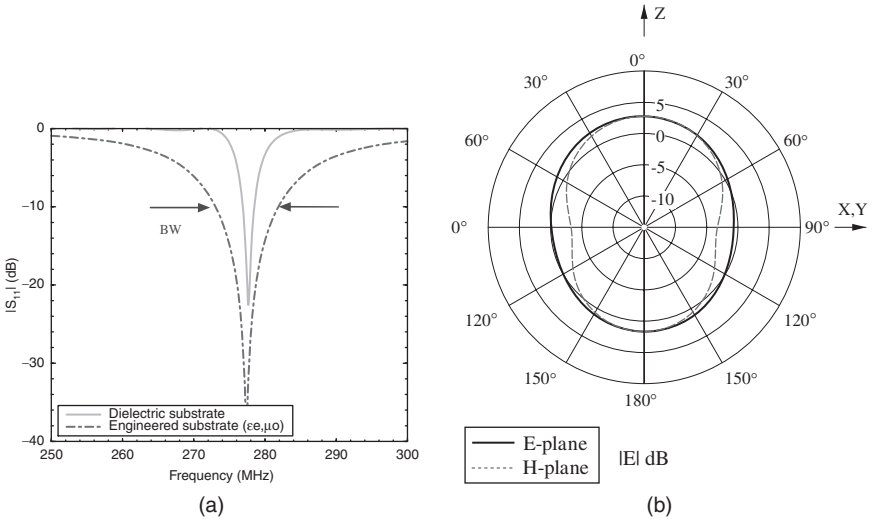


FIGURE 34-23 The performance of patch antenna printed on the magneto-dielectric substrate: (a) return loss and (b) radiation patterns (after Mosallaei and Sarabandi⁷⁰ © IEEE 2004)

As discussed above, the larger μ_r/ϵ_r provides the larger bandwidth. The layered design allows us to effectively accomplish a large μ_r/ϵ_r , although the hexaferrite itself has almost the same permittivity and permeability below the resonance. Since the patch antenna supports a dominant TEM wave with electric and magnetic fields polarized vertically and horizontally, respectively, they experience the effective dielectric and magnetic materials of $\epsilon_{er} = 3.84 - j0.004$ and $\mu_{er} = 8.61 - j0.16$.⁷⁰ This provides a miniaturization factor greater than 5 with considerably enhanced antenna bandwidth. The FDTD is applied to obtain the return loss and radiation patterns of antenna, as shown in Figure 34-23. The antenna resonance is at $f_0 = 277$ MHz and it provides a wide bandwidth of about $BW = 3.2\%$. The size of the antenna is around $0.09\lambda_0$ with a miniaturization factor of 5.4. The directivity of the antenna is $D_0 = 2.9$ dB and it has a front-to-back ratio 1.3 dB (ground plane size is $0.18\lambda_0 \times 0.18\lambda_0$). The calculated antenna efficiency is about $e_r = 67\%$. Notice that to achieve the same miniaturization factor utilizing only a dielectric material ($\mu_r = 1$) one must use $\epsilon_r = 23.7$. This reduces the bandwidth to about $BW = 0.5\%$ as shown in Figure 34-23a. The efficiency in this case for a dielectric loss tangent of 0.001 is about $e_r = 64\%$. Therefore, utilizing the magneto-dielectric substrate, one can offer a miniaturized wideband planar antenna with relatively high efficiency. The antenna bandwidth for the proposed magneto-dielectric substrate is about six times higher than that of the dielectric substrate. However, the major challenge is to practically achieve a low-loss magnetic material at any frequency of interest (e.g., GHz range).

Embedded-Circuit Metasubstrate

A metasubstrate constructed from loop circuits embedded in a low dielectric host medium can realize a magneto-dielectric substrate at any frequency of interest.⁷²⁻⁷⁵ Figure 34-24a shows the geometry of a periodic configuration of loop circuits. The building block unit cell of the structure is constructed from a metallic loop terminated to a large capacitor, modeling a resonant LC circuit. The loops are very small in size realizing artificial material molecules. They are oriented along the y-direction to offer the desired effective permeability

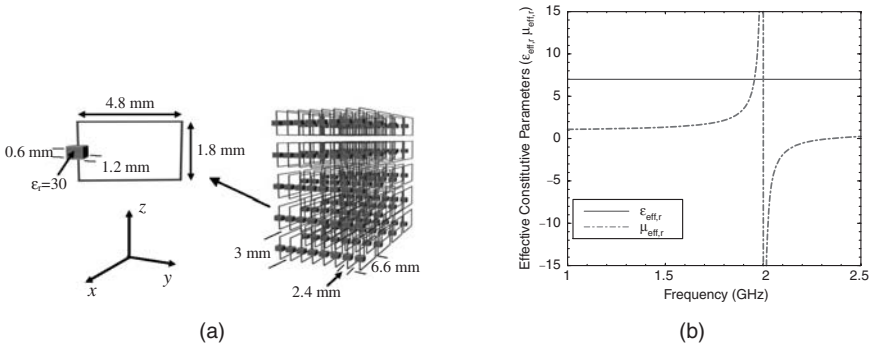


FIGURE 34-24 An embedded-circuit metamaterial with ϵ - μ parameters: (a) the geometry and (b) permeability performance (after H. Mosallaei and K. Sarabandi⁷⁵ © IEEE 2007)

in this direction. The gap coupling between the loops provides the required effective permittivity. The designed embedded-circuit medium has constant effective permittivity of $\epsilon_{\text{eff}} = 7.0\epsilon_0$, and the effective resonant permeability given by Mosallaei and Sarabandi⁷⁵ of

$$\mu_{\text{eff}} = \mu_0 \left(1 - (0.52)^2 \frac{1}{1 - (2.0)^2 / f^2 (\text{GHz})} \right) \quad (34-4)$$

These parameters are plotted in Figure 34-24b. Note that the resonant frequency of permeability, and thus the μ , can be properly tuned at any frequency of interest by simply changing the loop capacitors.

A metasubstrate is made from the embedded-circuit inclusions, as shown in Figure 34-25a. The operating mode of the patch antenna is a TM wave with electric and magnetic fields polarized along the z - and x -directions, respectively. Hence, the fields experience the effective parameters of the designed metamaterial. The FDTD is applied to characterize the antenna printed on the embedded-circuit metasubstrate, and the result for return loss is shown in Figure 34-25b. A resonant frequency of $f_0 = 1.74$ GHz (miniaturization factor of about 4) and bandwidth of about $BW = 1\%$ are determined. The patch antenna printed on a dielectric substrate with $\epsilon_r = 13.92$ (keeping the same resonant frequency) has bandwidth of about $BW = 0.6\%$. The metasubstrate offers a relatively wider bandwidth. It must be highlighted that the dispersion behavior of the permeability function plays an important role in degrading the impedance bandwidth performance. Both the dielectric and embedded-circuit substrates provide similar radiation patterns, as shown in Figure 34-26.

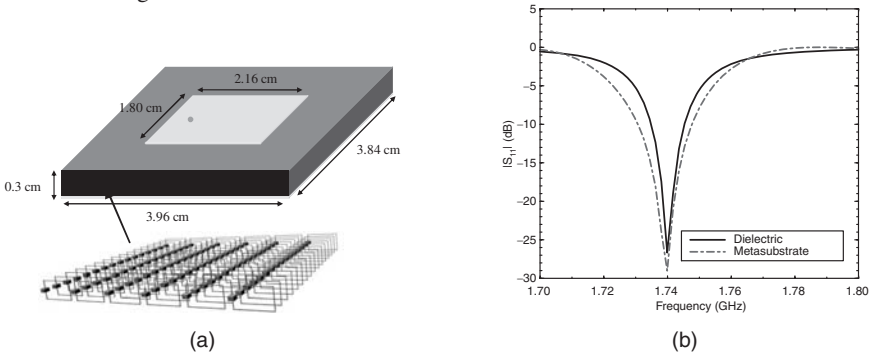


FIGURE 34-25 Patch antenna printed on an embedded-circuit metasubstrate: (a) the geometry and (b) return loss (after H. Mosallaei and K. Sarabandi⁷⁵ © IEEE 2007)

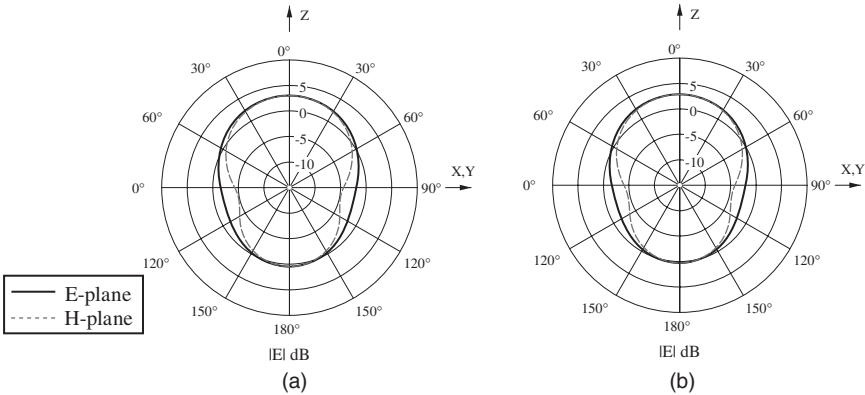


FIGURE 34-26 Radiation patterns of antenna printed on (a) dielectric and (b) metasubstrate (after H. Mosallaei and K. Sarabandi⁷⁵ © IEEE 2007)

To fabricate the metasubstrate, the first step is to realize an optimal design for the resonant loop circuits. A metallic loop terminated to an interdigitated capacitor seems like a reasonable choice for realizing LC inclusions. However, this is a nonoptimal use of the unit cell area because the interdigitation consumes too much valuable space inside the inductive loop and thereby inhibits optimal coupling to the incident magnetic field. It has been demonstrated in Buell et al⁷⁶ that a spiral loop is much more preferable because it uses less area to provide equivalent capacitance while simultaneously providing additional inductance, and hence additional permeability. A metasubstrate is fabricated utilizing the spiral inclusions illustrated in Figure 34-27.⁷⁶ It has a cell size of $\Delta x = \Delta z = 2$ cm, $\Delta y = 3.028$ mm (120 mils). The substrate is fabricated on 120-mil-thick Rogers RO-4003 dielectric. The spiral resonators are etched from 1/2 -oz-thick copper (0.017 mm) with a line-width (w) and spacing (s) of 0.127 mm (5 mils). In this design, $l_x = l_y = 16$ mm. To reduce substrate mass, 33/64-in-diameter air holes are drilled along the y -axis into the center of each spiral resonator cell. The final substrate mass is reduced by a factor of approximately 1/3, which is significant for a 2-cm-thick substrate. The substrate has a total size of $24 \times 24 \times 2$ cm and a weight of approximately 3.5 lb. It provides an effective permittivity of $\epsilon_{\text{reff}} = 13.13$,

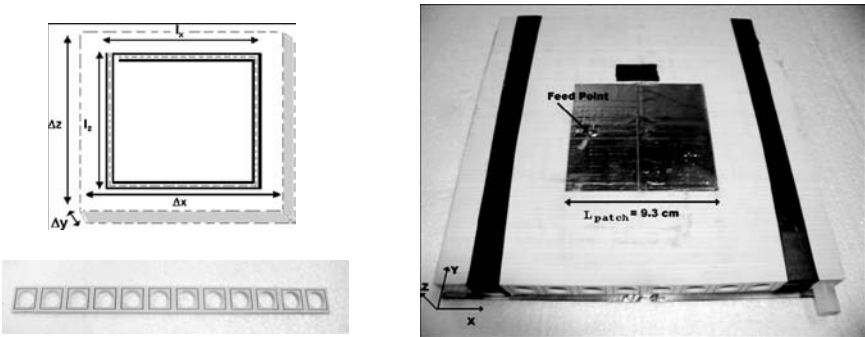


FIGURE 34-27 Fabricated metasubstrate utilizing spiral loop circuits (after Buell et al⁷⁶ © IEEE 2006)

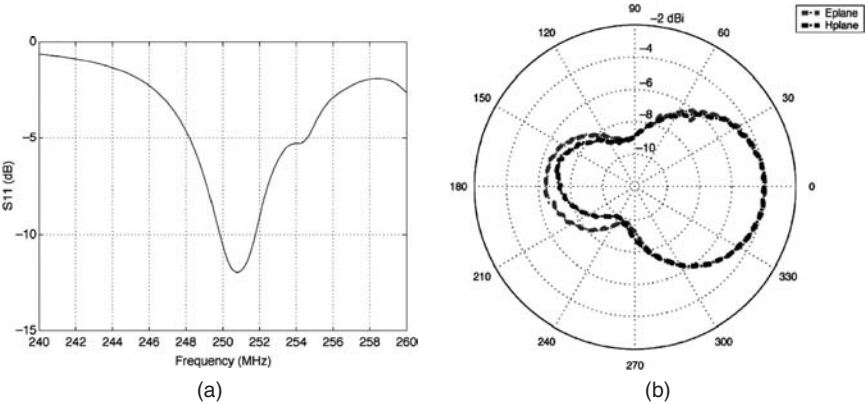


FIGURE 34-28 (a) Measured return loss and (b) radiation patterns of the patch antenna printed on the fabricated metasubstrate (after Buell et al⁷⁶ © IEEE 2006)

and an effective permeability with resonant frequency obtained around 285 MHz. Although here a design for the operation in UHF band is presented, the metamaterial unit cell can be properly redesigned for operation at any frequency of interest.

A probe-fed microstrip patch antenna resonant at 250 MHz is built on the metasubstrate (see Figure 34-27). To resonant at 250 MHz, the patch dimensions are found to be 9.3×9.3 cm. The measured return loss and radiation patterns are shown in Figure 34-28. A miniaturization factor of about 6.4 and efficiency around $e_r = 20\%$ are measured.

A better efficiency can be achieved if a thicker metallization is used. This will improve the magnetic loss tangent by a factor of 2. An additional source of loss is the dielectric loss tangent of the host dielectric material. Decreasing the host material dielectric loss tangent will significantly improve the efficiency of the dielectric medium, but not as strongly as improvements to the metallization would. To estimate the impact of changes in material loss tangent on antenna efficiency, the performance of a 250-MHz patch printed on a substrate with material parameters of $\epsilon_r = 9.8$, $\mu_r = 3.1$, and $\tan\delta = \tan\delta_c = \tan\delta_m$ is simulated in Figure 34-29. Obviously, the better material loss offers the better antenna radiation efficiency.

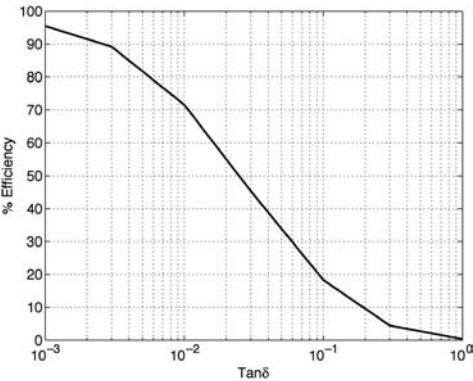


FIGURE 34-29 Radiation efficiency of a patch antenna at 250 MHz versus substrate material loss tangent on modeled metamaterial substrate (after Buell et al⁷⁶ © IEEE 2006)

In summary, it must be mentioned that although metasurfaces and metasubstrates theoretically present promising advantages; in practice, achieving low-loss wideband metamaterial would be a challenge. It is anticipated that the development of all-dielectric metamaterials to address some of these engineering concerns might help.

REFERENCES

1. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals* (Princeton, NJ: Princeton Univ. Press, 1995).
2. *IEEE Trans. Microwave Theory Tech., Special Issue on Electromagnetic Crystal Structures, Designs, Synthesis, and Applications*, vol. 47, no. 11 (November 1999).
3. *IEEE Trans. Antennas Propagat., Special Issue on Meta-materials*, vol. 51, no. 10 (October 2003).
4. *IEEE Trans. Antennas Propagat., Special Issue on Artificial Magnetic Conductors, Soft/Hard Surfaces and other Complex Surfaces*, vol. 53, no. 1 (January 2005).
5. Y. Rahmat-Samii and H. Mosallaei, "Electromagnetic Band-gap Structures: Classification, Characterization and Applications," *IEE-ICAP Symposium* (April 2001): 560–564.
6. B. A. Munk, *Frequency Selective Surfaces, Theory and Design* (New York: John Wiley & Sons, Inc., 2000).
7. B. A. Munk, *Finite Antenna Arrays and FSS* (New York: John Wiley & Sons, Inc. 2003).
8. V. G. Veselago, "The Electrodynamics of Substances with Simultaneous Negative Values of ϵ and μ ," *Sov. Phys. Usp.*, vol. 10, no. 4 (1996): 509–514.
9. J. B. Pendry, "Negative Refraction Makes a Perfect Lens," *Phys. Rev. Lett.*, vol. 85, no. 18 (October 2000): 3966–3969.
10. R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental Verification of a Negative Refractive Index of Refraction," *Science*, vol. 292 (April 2002): 77–79.
11. F. -R. Yang, K. -P. Ma, Y. Qian, and T. Itoh, "A Uniplanar Compact Photonic-Bandgap (UC-PBG) Structures and Its Applications for Microwave Circuit," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 8 (August 1999): 1509–1514.
12. D. Sievenpiper, L. Zhang, R. F. J. Broas, N. G. Alexopoulos, and E. Yablonovitch, "High-Impedance Electromagnetic Surfaces with a Forbidden Frequency Band," *IEEE Trans. Microwave Theory Tech.*, vol. 47, no. 11 (November 1999): 2059–2074.
13. A. S. Barlevy and Y. Rahmat-Samii, "Characterization of Electromagnetic Band-Gaps Composed of Multiple Periodic Tripods with Interconnecting Vias: Concept, Analysis, and Design," *IEEE Trans. Antennas Propagat.*, vol. 49, no. 3 (March 2001): 343–353.
14. P. -S. Kildal, "Artificial Soft and Hard Surfaces in Electromagnetics," *IEEE Trans. Antennas Propagat.*, vol. 38, no. 10 (October 1990): 1537–1544.
15. Rahmat-Sammi and F. Yang, "Development of Complex Artificial Ground Planes in Antenna Engineering," *Metamaterials: Physics and Engineering Explorations*, N. Engheta and R. Ziolkowski (eds.) (New York: John Wiley & Sons Inc., 2006).
16. M. Rahman and M. A. Stuchly, "Transmission Line-Periodic Circuit Representation of Planar Microwave Photonic Bandgap Structures," *Microwave Opt. Technol. Lett.*, vol. 30, no. 1 (July 2001): 15–19.
17. A. Taflov and S. Hagness, *Computational Electromagnetics: The Finite-Difference Time-Domain Method*, 2nd Ed. (Boston: Artech House, 2000).
18. M. A. Jensen and Y. Rahmat-Samii, "Performance Analysis of Antennas for Hand-held Transceiver Using FDTD," *IEEE Trans. Antennas Propagat.*, vol. 42 (August 1994): 1106–1113.
19. H. Mosallaei and Y. Rahmat-Samii, "Periodic Bandgap and Effective Dielectric Materials in Electromagnetics: Characterization and Applications in Nanocavities and Waveguides," *IEEE Trans. Antennas Propagat.*, vol. 51, no. 3 (March 2003): 549–563.
20. A. Aminian and Y. Rahmat-Samii, "Spectral FDTD: A Novel Computational Technique for the Analysis of Periodic Structures," 2004 IEEE APS International Symposium, Monterey, CA, June 20–26.

21. F. Yang and Y. Rahmat-Samii, "Reflection Phase Characterizations of the EBG Ground Plane for Low Profile Wire Antenna Applications," *IEEE Trans. Antennas Propagat.*, vol. 51 (October 2003): 2691–2703.
22. F. Yang and Y. Rahmat-Samii, "Polarization Dependent Electromagnetic Band-Gap Surfaces: Characterization, Designs, and Applications," *2003 IEEE AP-S Digest*, vol. 3 (June 2003): 339–342.
23. F. Yang and Y. Rahmat-Samii, "Polarization Dependent Electromagnetic Band Gap (PDEBG) Structures: Designs and Applications," *Microwave Opt. Technol. Lett.*, vol. 41, no. 6 (July 2004): 439–444.
24. Y. Q. Fu and N. C. Yuan, "Reflection Phase and Frequency Bandgap Characteristics of EBG Structures with Anisotropic Periodicity," *J. of Electromagnetics Waves and Applications*, vol. 19, no. 14 (2005): 1897–1905.
25. S. Clavijo, R. E. Diaz, and W. E. McKinzie III, "Design Methodology for Sievenpiper High-Impedance Surfaces: An Artificial Magnetic Conductor for Positive Gain Electrically Small Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 10 (October 2003): 2678–2690.
26. D. J. Kern et al, "The Design Synthesis of Multiband Artificial Magnetic Conductors Using High Impedance Frequency Selective Surfaces," *IEEE Trans. Antennas Propagat.*, vol. 53 (January 2005): 8–17.
27. M. F. Abedin and M. Ali, "Effects of EBG Reflection Phase Profiles on the Input Impedance and Bandwidth of Ultrathin Directional Dipoles," *IEEE Trans. Antennas Propagat.*, vol. 53 (November 2005): 3664–3672.
28. Z. Li and Y. Rahmat-Samii, "PBG, PMC and PEC Ground Planes: A Case Study for Dipole Antenna," *2000 IEEE APS International Symposium Digest*, vol. 4 (July 16–21, 2000): 2258–2261.
29. H. Mosallaei and K. Sarabandi, "Antenna Miniaturization and Bandwidth Enhancement Using a Reactive Impedance Substrate," *IEEE Trans. Antennas Propagat.*, vol. 52 (September 2004): 2403–2414.
30. H. Nakano, S. Okuzawa, K. Ohishi, H. Mimaki, and J. Yamauchi, "A Curl Antenna," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 11 (November 1993): 1570–1575.
31. F. Yang and Y. Rahmat-Samii, "A Low Profile Circularly Polarized Curl Antenna over Electromagnetic Band-Gap (EBG) Surface," *Microwave Opt. Technol. Lett.*, vol. 31, no. 3 (2001): 165–168.
32. H. Nakano et al, "Effects on the Radiation Characteristics of Using a Corrugated Reflector with a Helical Antenna and an Electromagnetic Band-Gap Reflector with a Spiral Antenna," *IEEE Trans. Antennas Propagat.*, vol. 53 (January 2005): 191–199.
33. F. Yang and Y. Rahmat-Samii, "A Low Profile Single Dipole Antenna Radiating Circularly Polarized Waves," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 9 (September 2005): 3083–3086.
34. F. Yang and Y. Rahmat-Samii, "Bent Monopole Antennas on EBG Ground Plane with Reconfigurable Radiation Patterns," *2004 IEEE APS International Symposium Digest*, vol. 2, (June 20–26, 2004): 1819–1822.
35. Y. Qian and T. Itoh, "Progress in Active Integrated Antennas and Their Applications," *IEEE Trans. Microwave Theory Tech.*, vol. 46, no. 11 (November 1998): 1891–1900.
36. J. T. Bernhard, "Reconfigurable Antennas and Apertures: State-of-the-Art and Future Outlook," *Proc. SPIE Conf. on Smart Electronics, MEMS, BioMEMS, and Nanotechnology*, vol. 5055 (March 2003): 1–9.
37. F. Yang and Y. Rahmat-Samii, "Patch Antennas with Switchable Slots (PASS) in Wireless Communications: Concepts, Designs, and Applications," *IEEE Antennas and Propagation Magazine*, vol. 47, no. 2 (April 2005): 13–29.
38. D. H. Schaubert, F. G. Farrar, A. Sindoris, and S. T. Hayes, "Microstrip Antennas with Frequency Agility and Polarization Diversity," *IEEE Trans. Antennas Propagat.*, vol. 29, no. 1 (January 1981): 118–123.
39. F. Yang and Y. Rahmat-Samii, "Patch Antenna with Switchable Slot (PASS): Dual Frequency Operation," *Microwave Optical and Technology Letters*, vol. 31, no. 3 (November 2001): 165–168.

40. M. Boti, L. Dussopt, and J. M. Laheurte, "Circularly Polarized Antenna with Switchable Polarization Sense," *Electronic Letters*, vol. 36, no. 18 (August 2000): 1518–1519.
41. F. Yang and Y. Rahmat-Samii, "A CP Polarization Diversity Patch Antenna Using Switchable Slots," *IEEE Microwave and Wireless Components Letters*, vol. 12, no. 3 (March 2002): 96–98.
42. W. E. McKinzie III and R. Fahr, "A Low Profile Polarization Diversity Antenna Built on an Artificial Magnetic Conductor," *2002 IEEE AP-s International Symposium Digest*, vol. 1 (June 2002): 762–765.
43. K. Chang, M. Li, T. -Y. Yun, and C. T. Rodenbeck, "Novel Low Cost Beam Steering Technique," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 5 (May 2002): 618–627.
44. K. W. Lee and R. G. Rojas, "Novel Scheme for Design of Adaptive Printed Antenna Element," *IEEE AP-S Digest*, vol. 3 (July 2000): 1252–1255.
45. D. Sievenpiper, J. Schaffner, R. Loo, G. Tansonan, S. Ontiveros, and R. Harold, "A Tunable Impedance Surface Performing as a Reconfigurable Beam Steering Reflector," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 3 (March 2003): 384–390.
46. J. Sor, C. -C. Chang, Y. Qian, and T. Itoh, "A Reconfigurable Leaky-Wave/Patch Microstrip Aperture for Phased-Array Applications," *IEEE Trans. Microwave Theory Tech.*, vol. 50 (August 2002): 1877–1884.
47. G. H. Huff, J. Feng, S. Zhang, and T. Bernhard, "A Novel Radiation Pattern and Frequency Reconfigurable Single Turn Square Spiral Microstrip Antenna," *IEEE Microwave and Wireless Component Letter*, vol. 13, no. 2 (February 2003): pp. 57–59.
48. G. P. Gauthier, A. Courty, and G. H. Rebeiz, "Microstrip Antennas on Synthesized Low Dielectric-Constant Substrate," *IEEE Trans. Antennas Propagat.*, vol. 45, no. 8 (August 1997): 1310–1314.
49. I. Papapolymerou, R. F. Drayton, and L. P. B. Katehi, "Micromachined Patch Antennas," *IEEE Trans. Antennas Propagat.*, vol. 46 (February 1998): 275–283.
50. J. S. Colburn and Y. Rahmat-Samii, "Patch Antennas on Externally Perforated High Dielectric Constant Substrates," *IEEE Trans. Antennas Propagat.*, vol. 47 (December 1999): 1785–1794.
51. D. R. Jackson, J. T. Williams, A. K. Bhattacharyya, R. L. Smith, S. J. Buchheit, and S. A. Long, "Microstrip Patch Designs that Do Not Excite Surface Waves," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 8 (August 1993): 1026–1037.
52. R. Gonzalo, P. de Maagt, and M. Sorolla, "Enhanced Patch Antenna Performance by Suppressing Surface Waves Using Photonic-Bandgap Substrates," *IEEE Trans. Microwave Theory Tech.*, vol. 47 (November 1999): 2131–2138.
53. R. Coccio, F. R. Yang, K. P. Ma, and T. Itoh, "Aperture Coupled Patch Antenna on UC-PBG Substrate," *IEEE Trans. Microwave Theory Tech.*, vol. 47 (Nov. 1999): 2123–2130.
54. F. Yang, C. -S. Kim, and Y. Rahmat-Samii, "Step-like Structure and EBG Structure to Improve the Performance of Patch Antennas on High Dielectric Substrate," *2001 IEEE AP-S Digest*, vol. 2 (July 2001): 482–485.
55. F. Yang and Y. Rahmat-Samii, "Microstrip Antennas Integrated with Electromagnetic Band-Gap (EBG) Structures: A Low Mutual Coupling Design for Array Applications," *IEEE Trans. Antennas Propagat.*, vol. 51, no. 10 (October 2003): 2936–2946.
56. Z. Iluz, R. Shavit, and R. Bauer, "Microstrip Antenna Phased Array with Electromagnetic Bandgap Substrate," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 6 (June 2004): 1446–1453.
57. L. Zhang, J. A. Castaneda, and N. G. Alexopoulos, "Scan Blindness Free Phased Array Design Using PBG Materials," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 8 (August 2004): 2000–2007.
58. L. Yang et al., "A Novel Compact Electromagnetic-Bandgap (EBG) Structure and Its Applications for Microwave Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 1 (January 2005): 183–190.
59. H. Boutayeb et al., "Analysis and Design of a Cylindrical EBG-based Directive Antenna," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 1 (January 2006): 211–219.
60. A. Neto et al., "On the Optimal Radiation Bandwidth of Printed Slot Antennas Surrounded by EBGs," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 4 (April 2006): 1074–1083.
61. Y. Q. Fu et al., "Mutual Coupling Reduction Between Large Antenna Arrays Using Electromagnetic Bandgap (EBG) Structures," *J. of Electromagnetic Waves and Applications*, vol. 20, no. 6 (2006): 819–825.

62. F. Yang, A. Aminian, and Y. Rahmat-Samii, "A Novel Surface Wave Antenna Design Using a Thin Periodically Loaded Ground Plane," *Microwave Optical and Technology Letters*, vol. 47, no. 3 (November 2005): 240–245.
63. Francis J. Zucker, *Antenna Engineering Handbook*, 3rd Ed., Richard C. Johnson (ed.), Chap. 12 (New York: McGraw-Hill, 1993).
64. R. Elliot, "Spherical Surface Wave Antennas," *IRE Trans. Antennas Propagat.*, vol. 4, no. 3 (July 1956): 422–428.
65. R. Hougardy and R. C. Hansen, "Scanning Surface Wave Antennas—Oblique Surface Waves over a Corrugated Conductor," *IRE Trans. Antennas Propagat.*, vol. 6, no. 4 (October 1958): 370–376.
66. L. B. Felson, "Radiation from a Tapered Surface Wave Antenna," *IRE Trans. Antennas Propagat.*, vol. 8 (November 1960): 577–586.
67. F. J. Zucker and J. A. Storm, "Experimental Resolution of Surface Wave Antenna Radiation into Feed and Terminal Patterns," *IEEE Trans. Antennas Propagat.*, vol. 18 (May 1970): 420–422.
68. Fan Yang, Yahya Rahmat-Samii, and Ahmed Kishk, "A Novel Surface Wave Antenna with a Monopole Type Pattern: A Thin Periodic Loaded Slab Excited by a Circular Disk," *2005 IEEE APS International Symposium Digest*, vol. 1A (July 3–8, 2005): 742–745.
69. D. R. Jackson and N. G. Alexopoulos, "Gain Enhancement Methods for Printed Circuit Antennas," *IEEE Trans. Antennas Propagat.*, vol. 33, no. 8 (September 1985): 976–987.
70. H. Mosallaei and K. Sarabandi, "Magneto-Dielectrics in Electromagnetics: Concept and Applications," *IEEE Trans. Antennas Propagat.*, vol. 52, no. 6 (June 2004): 1558–1567.
71. R. C. Hansen and M. Burke, "Antennas with Magneto-Dielectrics," *Microwave and Opt. Tech. Lett.*, vol. 26, no. 2 (July 2000): 75–78.
72. D. R. Smith, W. J. Padilla, D. C. Vier, S. C. A. Nemat-Nasser, and S. Schultz, "Composite Medium with Simultaneously Negative Permeability and Permittivity," *Phys. Rev. Lett.*, vol. 84, no. 18 (May 2000): 4184–4187.
73. H. Mosallaei, K. Sarabandi, and Y. Rahmat-Samii, "Novel Artificial Meta-Materials with Both ϵ - μ Parameters: A Composite Periodic Structure of Dielectric/Split Ring Resonators," *IEEE URSI International Symposium* (June 16–21, 2002): 40.
74. H. Mosallaei and K. Sarabandi, "Embedded-Circuit Meta-Materials for Novel Design of Tunable Electro-Ferromagnetic Permeability, Band-Gap, and Bi-anisotropic Media," *IEEE AP-S International Symposium and USNC/CNC/URSI National Radio Science Meeting* (June 22–27, 2003): 355–358.
75. ———, "Design and Modeling of Patch Antenna Printed on Magneto-Dielectric Embedded-Circuit Metasubstrate," *IEEE Trans. Antennas Propagat.*, vol. 55, no. 1 (January 2007): 45–52.
76. K. Buell, H. Mosallaei, and K. Sarabandi, "A Substrate for Small Patch Antennas Providing Tunable Miniaturization Factors," *IEEE Trans. Microwave Theory Tech.*, vol. 54, no. 1 (January 2006): 135–146.

Chapter 35

Reflectarray
Antennas

John Huang

*Jet Propulsion Laboratory
California Institute of Technology*

CONTENTS

35.1	INTRODUCTION.....	35-2
35.2	REVIEW OF DEVELOPMENT HISTORY	35-3
35.3	ANALYSIS AND DESIGN PROCEDURES	35-9
35.4	BANDWIDTH ISSUES	35-12
35.5	APPLICATIONS AND RECENT DEVELOPMENTS	35-14
35.6	SUMMARY	35-20

35.1 INTRODUCTION

This chapter gives an overview of the development history, key design methodologies, bandwidth issues, and applications for the reflectarray antenna, in particular the printed reflectarray.

The reflectarray is an antenna consisting of either a flat or slightly curved reflecting surface and an illuminating feed, as shown in Figure 35-1. On the reflecting surface, there are many isolated elements (e.g. open-ended waveguides, printed microstrip patches, dipoles, or rings) without any power-division transmission lines. The feed antenna spatially illuminates these isolated elements, which are predesigned to re-radiate and scatter the incident field with electrical phases that are required to form a planar phase front in the far-field distance. This operation is similar in concept to the use of a parabolic reflector that utilizes its unique curvature to reflect and form a planar phase front when a feed is placed at its focal point. Thus the term “flat reflector” is sometimes used to describe the reflectarray, which utilizes both technologies of parabolic reflector and array. As shown in Figure 35-2, there are several methods for reflectarray elements to achieve a planar phase front. For example, one is to use identical microstrip patches with different-length phase delay lines attached so that they can compensate for the phase delays over the different paths from the illuminating feed. Another is to use variable-size patches, dipoles, or rings so that elements can have different scattering impedances and, thus, different phases to compensate for the different feedpath delays. With the third method, for circular polarization only, the reflectarray has all identical circularly polarized elements but with different angular rotations to compensate for the feedpath length differences.

Reflectarrays using printed microstrip elements have been developed to achieve low reflecting surface profile, small antenna mass, and low manufacturing cost. These reflectarrays combine some of the salient features of the traditional parabolic reflector antenna and the microstrip array technology. Similar to a parabolic reflector, the reflectarray can achieve very good efficiency (> 50 percent) for a very large aperture since no power divider is needed and thus very little resistive insertion loss is encountered here. On the other hand, very similar to an array antenna, the reflectarray can have its main beam designed to tilt at a large angle ($> 50^\circ$) from its broadside direction. Low-loss electronic phase shifters can be implanted into the elements for wide-angle electronic beam scanning. With this beam scanning capability

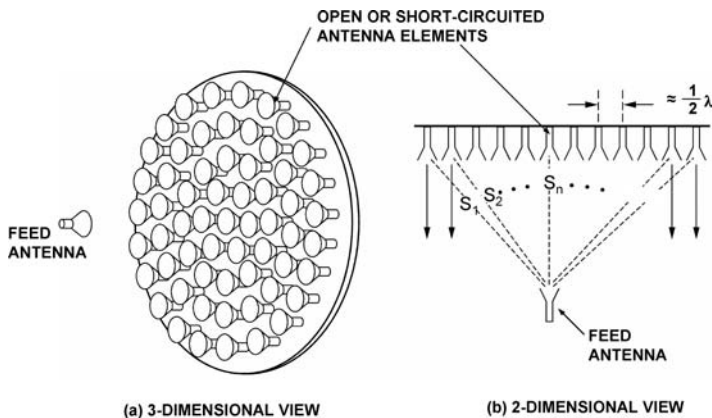


FIGURE 35-1 Configuration of a reflectarray antenna

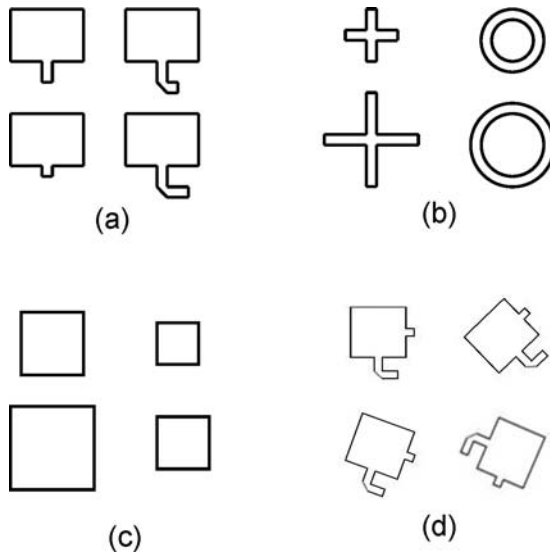


FIGURE 35-2 Various reflectarray elements: (a) identical patches with variable-length phase delay lines, (b) variable-size dipoles or loops, (c) variable-size patches, and (d) variable angular rotations

of the reflectarray, the complicated high-loss beamforming network and high-cost transmit/receive (T/R) amplifier modules of a conventional phased array are no longer needed.

One significant advantage of the printed reflectarray is that, when a large aperture (e.g. 10-m size) spacecraft antenna requires a deployment mechanism, the flat structure of the reflectarray allows a much simpler and more reliable folding or inflation mechanism to be accomplished than is possible with a specifically curved surface of a parabolic reflector. The flat reflecting surface of the reflectarray also lends itself to flush mounting onto an existing flat structure without adding a significant amount of mass and volume to the overall system structure. A reflectarray with hundreds or thousands of elements, in the form of a printed microstrip antenna, can be fabricated with a simple and low-cost etching process, especially when produced in large quantities. Another major feature of this antenna is that, with a large number of elements in a reflectarray having elemental phase adjustment capability, it can achieve a very accurate contour beam shape with a phase synthesis technique.

With all the above capabilities, there is one distinct disadvantage associated with the reflectarray antenna: its inherent narrow bandwidth, which generally cannot exceed much beyond 10 percent depending on its element design, aperture size, focal length, etc. This narrow bandwidth behavior will be discussed further in a later section. Although the reflectarray has narrow bandwidth, due to its multitude of capabilities, the development, research, and application of the printed reflectarray antenna will continue throughout this century and beyond.

35.2 REVIEW OF DEVELOPMENT HISTORY

Although the reflectarray antenna has not been widely exposed to the antenna community until recently, it was invented more than 40 years ago. Since then, many interesting reflectarray technologies have been developed. This section briefly reviews the development of

this antenna technology throughout the past several decades. This review will not only inform a reader about this technology's evolution, but also may lead him or her, by combining old thoughts with new technologies, to other new discoveries.

Waveguide Reflectarray in the 1960s

The reflectarray antenna concept was first conceived in the early 1960s.¹ Either open- or short-ended waveguide elements with variable-length waveguides, as illustrated in Figure 35-3, were used to demonstrate the capability of achieving co-phased reradiated far-field beams. Because most wireless operations during this early time were done at relatively low microwave frequencies, the large-waveguide reflectarrays resulted in very bulky and heavy antennas, and thus this antenna concept was not pursued until more than ten years later. In addition, the efficiencies of these reflectarrays were not studied and optimized.

Spiralphase Reflectarray in the 1970s

In the mid 1970s, a very clever concept of “spiraphase” reflectarray was developed,² where switching diodes, as illustrated in Figure 35-4, were used in a four-arm spiral or dipole element of a circularly polarized reflectarray to electronically scan its main beam to large angles from the broadside direction. This is possible because, by angularly rotating a circularly polarized radiating element, its propagating electrical phase will also change by an appropriate amount proportional to the amount of rotation. However, due to the thick spiral cavity (quarter-wavelength depth) and large electronic components, the spiraphase reflectarray was still relatively bulky and heavy. Its aperture efficiency was still relatively poor. Thus, no continued development effort was followed. It should be noted here that, in order to have good efficiency for the reflectarray, the intricate relations between the element beamwidth, element spacing, and focal-length/diameter (f/D) ratio must be well designed; otherwise, a large backscattered component field or a mismatched surface impedance would result.

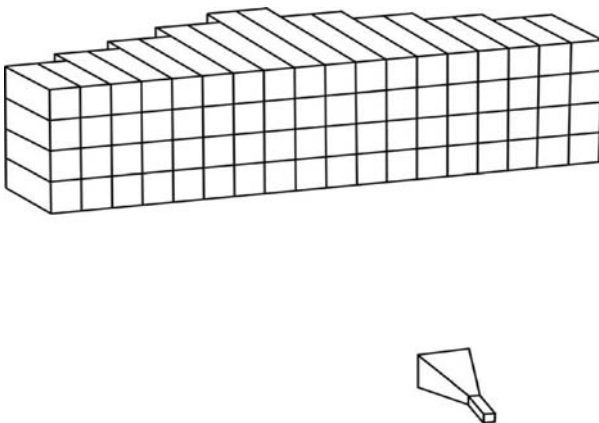


FIGURE 35-3 Conceptual drawing of the very early reflectarray using open-ended waveguides as elements



FIGURE 35-4 Reflectarray using a four-arm spiral element with switching diodes at center to achieve a 2-bit phase-shift system for circular polarization

Microstrip Reflectarray in the 1980s

With the introduction of printable microstrip antennas, the technologies of reflectarray and microstrip radiators were combined, a typical configuration of which is illustrated in Figure 35-5. The first mention of using microstrip elements for reflectarray³ was in 1978. During the same year, the first attempt to analyze the microstrip reflectarray element using the infinite array approach was carried out.⁴ Since then, various printed microstrip reflectarray antennas were developed in the late 1980s and early 1990s for the purpose of achieving reduced antenna size and mass. These printed reflectarrays came in various forms, as shown in Figure 35-2, and all had flat, low-profile, low-mass reflecting surfaces. The ones that used identical patch elements with different-length phase delay lines⁵⁻¹⁰ had their elements arranged similar to those shown in Figure 35-2a. The phase delay lines, having lengths on the order of half-wavelength long or less, were used to compensate for the phase

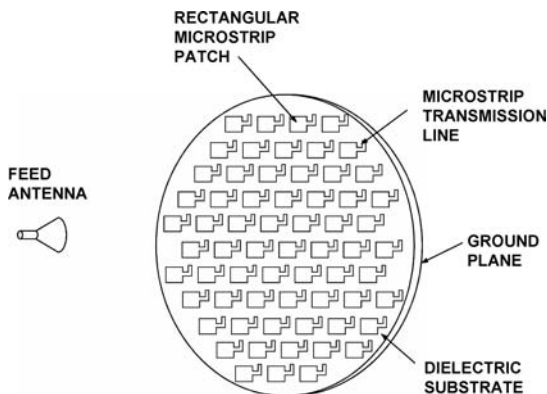


FIGURE 35-5 Microstrip reflectarray with identical patches but different-length phase delay lines

differences of different path lengths from the illuminating feed. The second approach, shown in Figure 35-2*b*, used elements that are made of printed dipoles with variable dipole lengths.¹¹ Different dipole lengths will yield different scattering impedances, which then provide the different phases needed to compensate for the different path-length delays. Similarly, microstrip patches with variable patch sizes,¹² shown in Figure 35-2*c*, were also developed. The concept of using circularly polarized microstrip patches with identical size but variable angular rotations,^{13,14} shown in Figure 35-2*d*, to form a co-phased far-field reflectarray beam was conceived and developed.

Recent Developments

In addition to the various reflectarray elements shown in Figure 35-2, several other reflectarray or equivalent developments during the 1990s are worth mentioning here. Printed variable-length dipole elements, shown in Figure 35-6, were used to form a frequency-scanned grating-reflector antenna with an offset feed.¹⁵ Printed annular rings of variable diameters arranged in Fresnel Zone configuration, as shown in Figure 35-7, were also used to focus the beam.¹⁶

In the 1996 Phased Array Conference, a 94-GHz monolithic reflectarray¹⁷ fabricated in a single wafer, using 1-bit PIN diode phase shifters, was reported to achieve wide-angle ($\pm 45^\circ$) electronic beam scanning. In the same conference, a 35-GHz reflectarray, using waveguide/dielectric elements with 3-bit ferrite phase shifters,¹⁸ was also reported to achieve $\pm 25^\circ$ beam scanning. One proposed technique,¹³ although not yet developed, is worth mentioning here. By using the angular rotation technique with circularly polarized elements, as depicted in Figure 35-8, miniature or micro-machined motors could be placed under each element to achieve wide-angle beam scanning without the need of T/R modules and phase shifters. For application in the spacecraft area, a deployable and low-mass

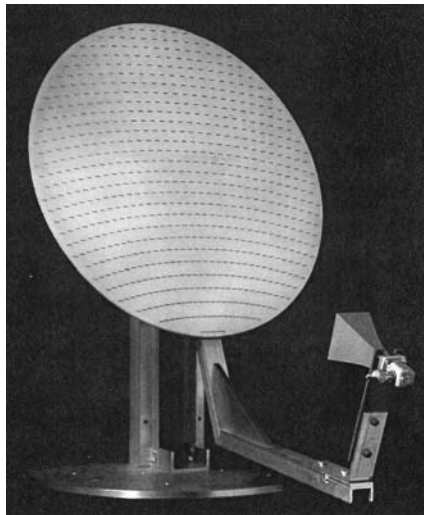


FIGURE 35-6 Printed dipole frequency-scanning grating-reflector antenna

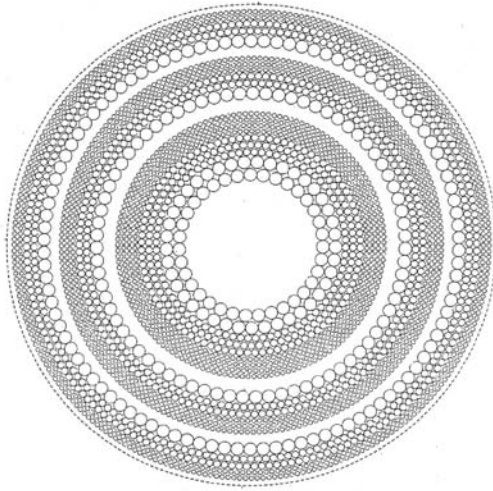


FIGURE 35-7 Printed Fresnel Zone antenna with variable-diameter annular rings

1-m-diameter inflatable reflectarray antenna¹⁹ at the X-band frequency was developed. Another unique spacecraft application of the reflectarray was conceived²⁰ and developed²¹ by using its many elements, with a numerical phase synthesis technique, to form a uniquely shaped contour beam.

From all the above developments, it can be seen that, at the end of the 20th century, the reflectarray antenna technology was becoming mature enough for possible applications throughout the microwave and millimeter-wave spectra.

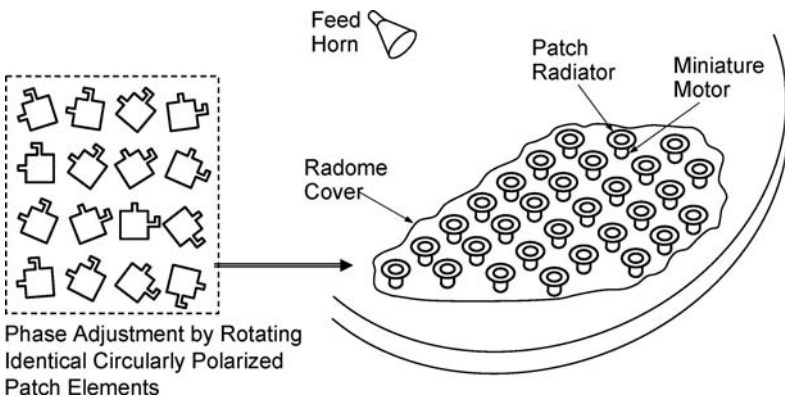


FIGURE 35-8 Miniature micro-machined motors used to achieve beam scan by employing the angular rotation technique with CP elements

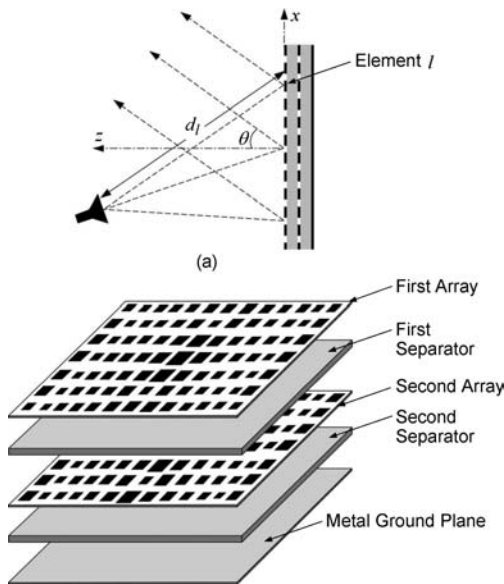


FIGURE 35-9 Multilayer reflectarray to achieve wider bandwidth

In the early 2000s, the development of reflectarray has mushroomed, and several performance improvement techniques are worth mentioning here. One uses multilayer stacked patches, as shown in Figure 35-9, to improve the reflectarray bandwidth from a few percent to more than 10 percent.²² As an extension to the 1-m X-band inflatable reflectarray mentioned above, a 3-m Ka-band inflatable reflectarray consisting of 200,000 elements was also developed,²³ which currently is the electrically largest reflectarray. An X-band amplifying reflectarray was developed²⁴ for each element of the reflectarray, as shown in Figure 35-10, to amplify the transmitted signal and, thus, achieve very high overall radiated power. To achieve

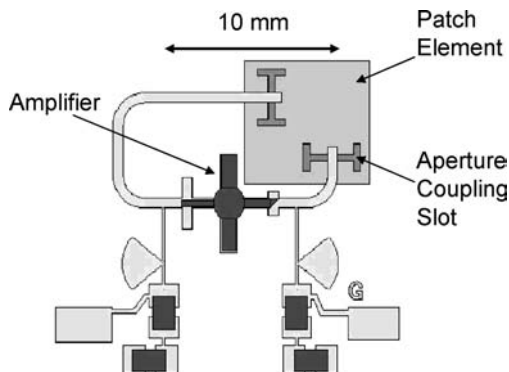


FIGURE 35-10 Unit-cell element of an X-band amplifying reflectarray

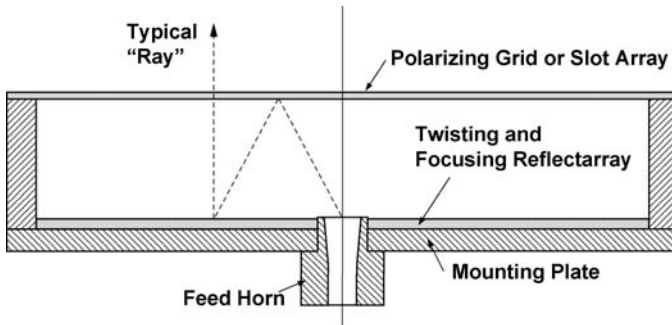


FIGURE 35-11 Folded reflectarray to achieve a more compact antenna profile

good antenna efficiency, the most critical segment of the reflectarray design is its elements. The element performance was optimized by using the technique of genetic algorithm.²⁵ The reflectarray using subreflector and array feed configuration to achieve fine beam scanning was also studied.²⁶ To combat the shortcoming of narrow bandwidth, dual-band multilayer reflectarrays using annular rings²⁷ and crossed dipoles²⁸ are also being developed. Another development that is worth mentioning here is a folded reflectarray configuration,²⁹ where two reflecting surfaces, as depicted in Figure 35-11, are used to reduce the overall antenna profile due to feed height of a conventional reflectarray.

35.3 ANALYSIS AND DESIGN PROCEDURES

The design and analysis of a reflectarray can be separated into four essential steps, which are separately discussed below.

Element Characterization

The most important and critical segment of the reflectarray design is its element characterization. If the element design is not optimized, the reflectarray will not scatter the signal from the feed effectively to form an efficient far-field beam. Its beamwidth must correlate correctly with the reflectarray's f/D ratio to accommodate all incident angles from the feed. Its phase change versus element change (patch size, delay line length, etc.) must be calibrated correctly. One of the most popular techniques for calibrating the phase is to use the infinite array approach^{12,30} to include local mutual coupling effect due to surrounding elements. It is not feasible for the current computer technology to have a complete rigorous solution to include all the mutual coupling effect of all different elements since the reflectarray generally consists of too many elements. The infinite array approach, which assumes all elements are identical, can be done by using the method of moments (MOM) technique^{12,30} or equivalently by using a finite difference time domain (FDTD) analysis on a unit cell of a single element.³¹ A mathematical waveguide simulator, which simulates the infinite array approach, can also be adapted by using the commercial software—HFSS (a finite element technique)—to achieve the element phase information.

All of these techniques are used to derive the phase-versus-element-change curve, which is generally an S -shaped curve with a nonlinear relationship, as illustrated in Figure 35-12.

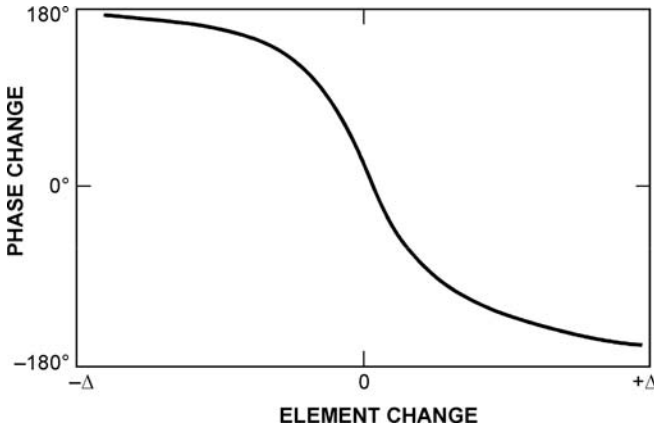


FIGURE 35-12 A typical *S* curve of a reflectarray element phase change versus element change

The antenna designer should minimize the slope at the center of the curve so that the phase change will not be overly sensitive to the element change. If the curve is too steep, the element change or fabrication tolerance may become an issue, in particular at high microwave frequencies.

Phase Delay Calculation

The path lengths from the feed to all elements are all different, which leads to different phase delays. To compensate for these phase delays, the elements must have corresponding phase advancements designed according to a unique *S* curve similar to that shown in Figure 35-12. The following gives an example of how the compensating phase is calculated for each element of a reflectarray with a broadside-directed beam. The differential path length for each element is given as

$$\Delta L_{m,n} = L_{m,n} - L_{o,o} \quad (35-1)$$

where $L_{m,n}$ is the distance between the feed and the m nth element, which can be obtained by a simple geometry calculation. $L_{o,o}$ is the distance between the feed and a reference point on the reflectarray surface, e.g. the center point. $\Delta L_{m,n}$ is thus the differential feedpath length for the m nth element. To achieve a collimated radiation, the phase advancement $\Delta\Phi_{mn}$ needed for the m nth element is given by

$$\Delta\Phi_{mn} \text{ in degrees} = [\Delta L_{m,n} / \lambda_o - \text{integer of } (\Delta L_{m,n} / \lambda_o)] \times 360. \quad (35-2)$$

The above indicates that the compensating phase can be repeated every 360° and that the portion that is an integer multiple of a wavelength, or 360° , can be deleted.

Pattern Calculation

With all elements' compensating phases known, the far-field radiation patterns can be calculated by the conventional array theory, where the radiation of all elements is summed

together as follows. Consider a planar array consisting of $M \times N$ elements that are nonuniformly illuminated by a low-gain feed at position vector \vec{r}_f . Let the desired beam direction be specified by unit vector \hat{u}_o . Then the far-field pattern of the reflectarray in the \hat{u} direction will be of the form

$$E(\hat{u}) = \sum_{m=1}^M \sum_{n=1}^N F(\vec{r}_{mn} \bullet \vec{r}_f) \cdot A(\vec{r}_{mn} \bullet \hat{u}_o) \cdot A(\hat{u} \bullet \hat{u}_o) \cdot \exp[jk(|\vec{r}_{mn} - \vec{r}_f| + \vec{r}_{mn} \bullet \hat{u}) + j\alpha_{mn}] \quad (35-3)$$

where F is the feed pattern function, A is the reflectarray element pattern function, \vec{r}_{mn} is the position vector of the m nth element, and α_{mn} is the required compensating phase of the m nth element calculated by Eq. 35-2. $\cos^q \theta$ factor is used for both F and A functions with no azimuth (ϕ) dependence.

Geometry Design

To determine the geometry of a reflectarray is basically to determine its f/D ratio, which is governed by its desired aperture efficiency. The aperture efficiency (η_a) can be defined as the product of the illumination (η_i) and spillover (η_s) efficiencies: $\eta_a = \eta_i \times \eta_s$. By integrating the pattern function of Eq. 35-3, the illumination efficiency for a center-fed reflectarray can be obtained in a close form³² as given by

$$\eta_i = \frac{\left[\left((1 + \cos^{q+1} \theta_e) / (q+1) \right) + \left((1 - \cos^q \theta_e) / q \right) \right]^2}{2 \tan^2 \theta_e \left[(1 - \cos^{2q+1} \theta_e) / (2q+1) \right]} \quad (35-4)$$

and the spillover efficiency is given by

$$\eta_s = 1 - \cos^{2q+1} \theta_e \quad (35-5)$$

where q is the exponent of the feed pattern function represented by $\cos^q \theta$, and θ_e is half of the subtend angle from the feed to the reflectarray aperture. The reflectarray element is approximated by a cosine function. Equations 35-4 and 35-5 are calculated by assuming a circular aperture only for the demonstration of the design procedures. Similar closed-form equations can be easily obtained for square, rectangular, or elliptical apertures by performing proper integrations. To give an example of how Equations 35-4 and 35-5 can be utilized to optimize a reflectarray design, Figure 35-13 shows the calculated curve of spillover and illumination efficiencies versus the feed pattern factor q (feed beamwidth) for a 0.5-m 32-GHz reflectarray with a fixed f/D ratio of 1.0 ($\theta_e = 26.6^\circ$). It demonstrates that the maximum aperture efficiency is achieved at $q = 10.5$ or when the feed has a -3 -dB beamwidth of 29° . Another curve, shown in Figure 35-14, gives aperture efficiency as a function of f/D ratio for the same 0.5-m 32-GHz reflectarray when the feed beamwidth is fixed at 33.4° with $q = 8$. In this case, the maximum aperture efficiency is achieved when the f/D ratio is 0.87. It can be seen that curves derived from Eqs. 35-4 and 35-5 are essential in obtaining an optimum efficiency design.

The above discussion has been limited to center-fed reflectarray. Offset-fed reflectarray can also be optimally designed by using equations similar to Eqs. 35-4 and 35-5.

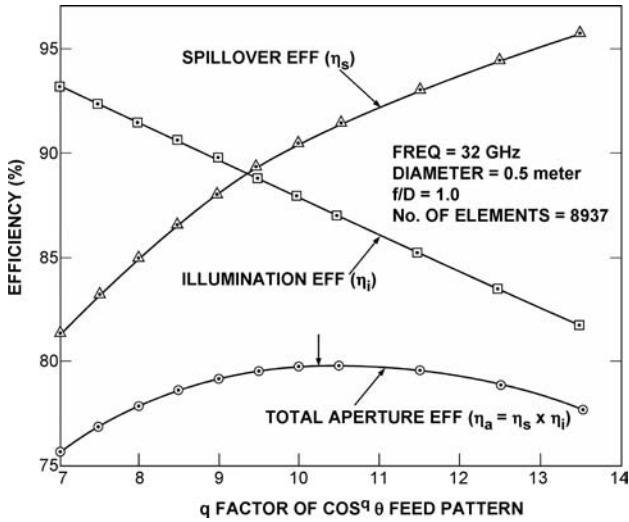


FIGURE 35-13 Spillover and illumination efficiencies versus feed pattern shape

35.4 BANDWIDTH ISSUES

The bandwidth performance of a reflectarray¹³ is no match to that of a parabolic reflector, where, theoretically, infinite bandwidth exists. For a printed microstrip reflectarray, its bandwidth is primarily limited by two factors. One is the narrow bandwidth of the microstrip patch elements on the reflectarray surface, and the other is the differential spatial phase delay.

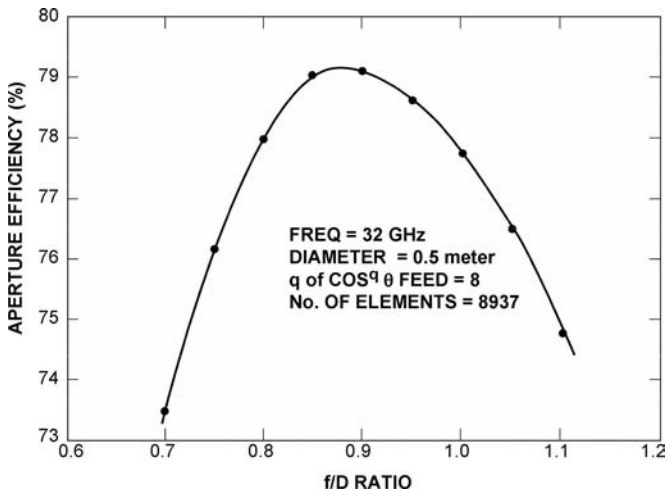


FIGURE 35-14 Aperture efficiency versus f/D ratio

Limit by Element Bandwidth

The microstrip patch element generally has a bandwidth of about 3 to 5 percent. To achieve wider bandwidth for a conventional microstrip array, techniques such as using thick substrate for the patch, stacking multiple patches,²² and using sequentially rotated subarray elements have been employed. Bandwidths of more than 15 percent have been reported.

Limit by Differential Spatial Delay

The second reflectarray limiting factor, the differential spatial phase delay, can be best explained by referring to Figure 35-15, where the differential spatial phase delay, ΔS , is the phase difference between the two paths S1 and S2 from the feed to the reflectarray elements. This ΔS can be many multiples of the wavelength (λ) at the center operating frequency. It can be expressed as $\Delta S = (N + d)\lambda$ where N is an integer and d is a fractional number of a free-space wavelength λ . At each element location, d is compensated by an appropriate phase delay achieved by the reflectarray element design (achieved by variable patch size, variable phase delay line length, etc.). As frequency changes, the factor $(N + d)\lambda$ becomes $(N + d)(\lambda + \Delta\lambda)$. Since the design and the compensating phase for each element are fixed for the center frequency, a frequency excursion error will occur in the reradiated phase front. The amount of phase change in each path when compared to a reference path, say S1, is $(N + d)\Delta\lambda$, which can be a significant portion of a wavelength (360°).

To reduce the amount of frequency excursion error, the integer number N must be reduced. There are several methods to reduce N . One is to design the reflectarray with a larger f/D ratio and hence to minimize the difference between paths S1 and S2. The second way is simply to avoid the use of a reflectarray with a large electrical diameter. The effects of f/D ratio and diameter on bandwidth performance were given previously in Figures 35-13 and 35-14. The third method to reduce frequency excursion error is to use time delay lines or partial time delay lines instead of the phase delays. In other words, when using the phase delay line technique (not the variable patch size technique), instead of using $d\Delta\lambda$ for the delay line length, $(N + d)\Delta\lambda$ could be used for the delay line. Certainly, additional line insertion loss and needed real estate for the lines are issues to be encountered.

Another method to increase the bandwidth is to use, instead of a complete flat reflectarray surface, a concavely curved reflectarray with piecewise flat surfaces. This curved

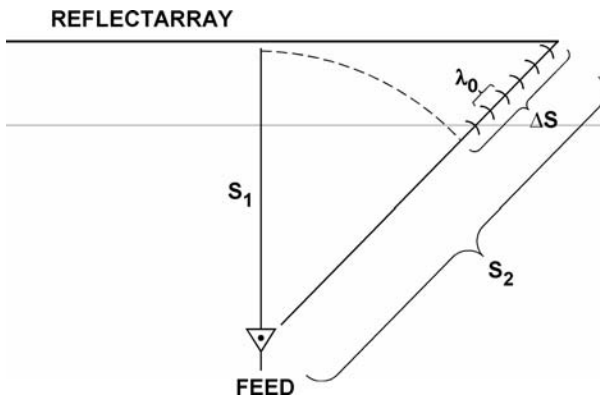


FIGURE 35-15 Differential spatial phase delay of reflectarray

reflectarray will continue to have advantages over a completely curved parabolic reflector; for example, its beam is able to be scanned to large angles with a phase shifter inserted into each element, and, for a space-deployable antenna, the piecewise flat surfaces in some cases are more easily folded into a smaller stowed volume. To mitigate the bandwidth problem, a recent technique of using multilayer stacked-patch elements²² not only has increased the element bandwidth but also has reduced the effect of differential spatial phase delay. As a net result, the bandwidth has increased from a few percent to more than 10 percent. Multiband techniques can also be applied to the reflectarray. Recently, two dual-band techniques have been developed for the X- and Ka-band frequencies. One uses double-layer membranes with two different-size rings and variable angular rotations,²⁷ and the other also uses a double-layer approach with X-band crossed dipoles over Ka-band patches.²⁸ To summarize, although the narrow bandwidth characteristic is the primary shortcoming of a reflectarray, there are several techniques that can be employed to alleviate the problem associated with the bandwidth issue.

35.5 APPLICATIONS AND RECENT DEVELOPMENTS

In addition to the possible reflectarray applications mentioned in the introduction and review sections, it is worthwhile here to present some details of several important applications and recent developments.

Inflatable Reflectarray

A Ka-band circularly polarized inflatable reflectarray²³ with a 3-m-diameter aperture was developed by the Jet Propulsion Laboratory (JPL) and ILC Dover, Inc. for NASA's future spacecraft communication antenna application. As shown in Figure 35-16, the antenna

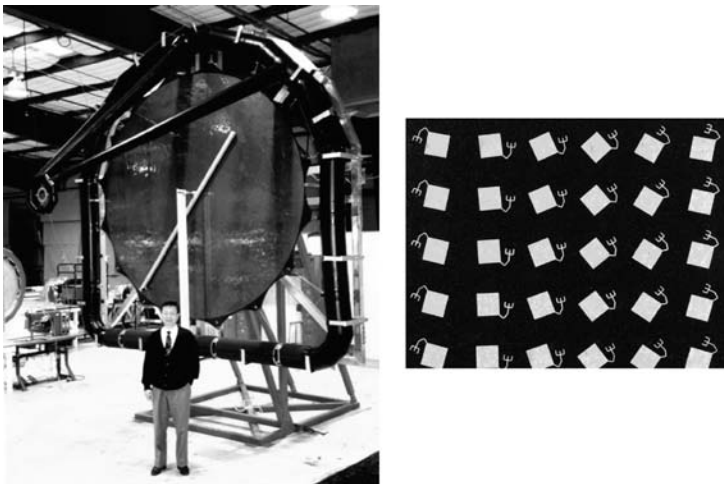


FIGURE 35-16 A 3-m Ka-band inflatable reflectarray (the shining structure in front of the aperture is the surface flatness measurement device). The second photo shows the expanded view of reflectarray elements on thin membrane.

uses a torus-shaped inflatable tube to support and tension a 3-m-thin membrane reflectarray surface. This circularly polarized reflectarray, having approximately 200,000 elements using the variable angular rotation technique,^{13,14} is considered to be electrically the largest reflectarray ever built. It is much easier for the reflectarray, having a “natural” flat surface, to maintain its required surface tolerance (0.3-mm rms in this case) by the inflatable structure than a “nonnatural” parabolic surface; in particular, for long space flight. This inflatable antenna was later improved to equip it with rigidizable inflatable tubes^{23,33} in order to survive the hazardous space environment, such as bombardment by space debris and strenuous thermal effect. This reflectarray achieved an aperture efficiency of 30 percent with room for improvement and excellent far-field pattern shape with average sidelobe and cross-polarization levels below -40 dB, as shown in Figure 35-17.

Shaped Contour Beam

A second important development of the reflectarray is the achievement of a shaped contour beam by using phase synthesis technique. This reflectarray, shown in Figure 35-18, was developed by the University of Massachusetts²¹ for a commercial application to provide Earth contour beam coverage. A typical calculated contour beam of this antenna, by using phase synthesis technique, is given in Figure 35-19. Since a reflectarray generally has more than thousands of elements, it thus has many degrees of freedom in design to provide an accurate and uniquely required contour beam. However, this capability of beam shaping is limited by its frequency bandwidth. A more recent reflectarray development³⁴ using triple-layer stacked patches achieved a specified contour beam shape within a relatively wide bandwidth of 10 percent.

Dual-band Cassgrain Reflectarray

A third important development is a dual-frequency reflectarray, where the two frequencies are widely separated, such as the X- and Ka-bands. The developed prototype antenna,

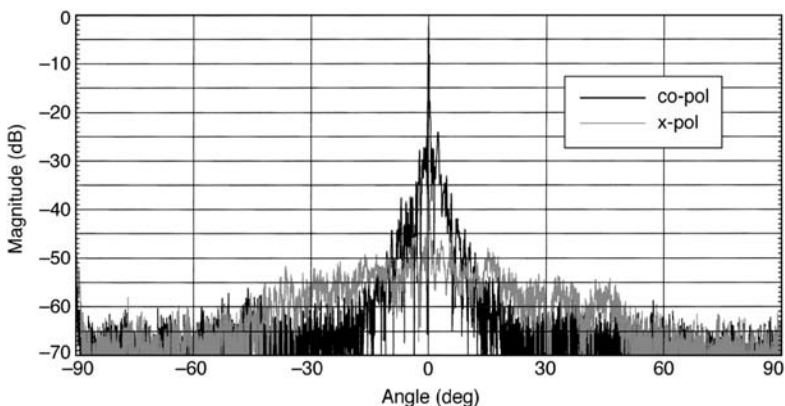


FIGURE 35-17 Measured radiation pattern of the 3-meter Ka-band inflatable reflectarray



FIGURE 35-18 Ku-band reflectarray with shaped contour beam capability (*Courtesy of Professor Dave Pozar, Univ. of Massachusetts*)

shown in Figures 35-20 and 35-21, is circularly polarized and uses variable angularly rotated annular rings.^{14,27} It was developed by Texas A&M University for JPL/NASA's future space communication application. This antenna, with a Cassegrain offset feed and an aperture diameter of 0.75 m, uses a multilayer technique where the X-band annular rings

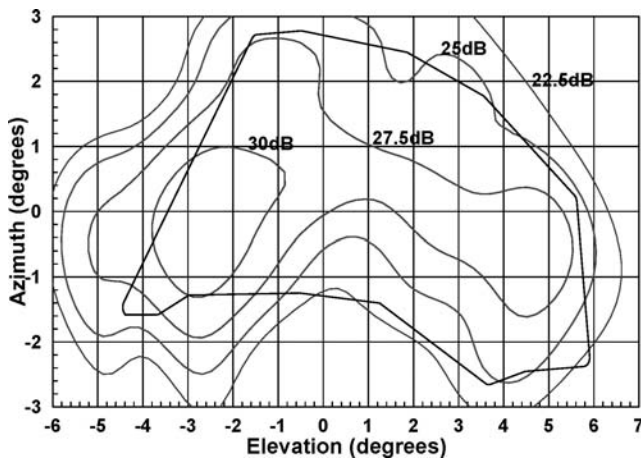


FIGURE 35-19 A measured contour beam plot of the reflectarray shown in Figure 35-18



FIGURE 35-20 X/Ka dual-band Cassegrain reflectarray antenna (0.75-m diameter) using annular ring elements

are placed above the Ka-band rings and serve as a frequency-selective surface to let the Ka-band signal pass through. The measured results indicate that there is very little impact on the X-band performance due to the presence of the Ka-band elements. The measured radiation patterns at the X-band and Ka-band frequencies are shown in Figures 35-22 and 35-23, respectively. These patterns show excellent behavior with relatively low sidelobe and low cross-polarization levels. However, the measured Ka-band gain of the dual-frequency dual-layer antenna is about 1.0 dB lower than the Ka-band-alone antenna. The Ka-band-alone reflectarray has a measured aperture efficiency of 50 percent, while the dual-frequency dual-layer antenna has a Ka-band efficiency of about 40 percent. In other words, the X-band annular rings did impact somewhat the Ka-band performance. Efforts need to be carried out in the future to minimize this impact.

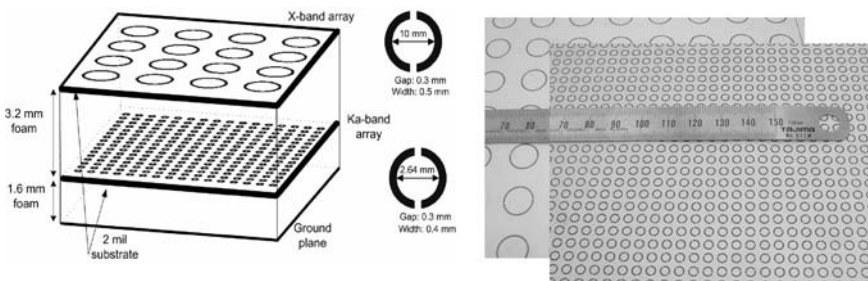


FIGURE 35-21 Sketch and photo of the X/Ka dual-band reflectarray antenna showing two membranes with annular ring elements

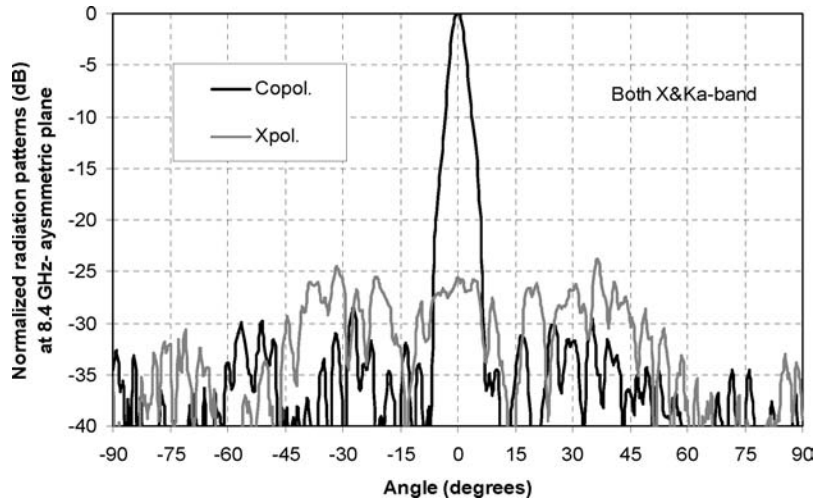


FIGURE 35-22 Measured X-band radiation pattern of the X/Ka dual-band reflectarray

Foldable Piecewise-Flat Globally Curved Reflectarray

A recent development that is worth mentioning is a reflectarray having a rectangular aperture that is intended for NASA/JPL’s Wide Swath Ocean Altimeter (WSOA) radar application. This reflectarray uses variable-size patches as elements. The required rectangular aperture,

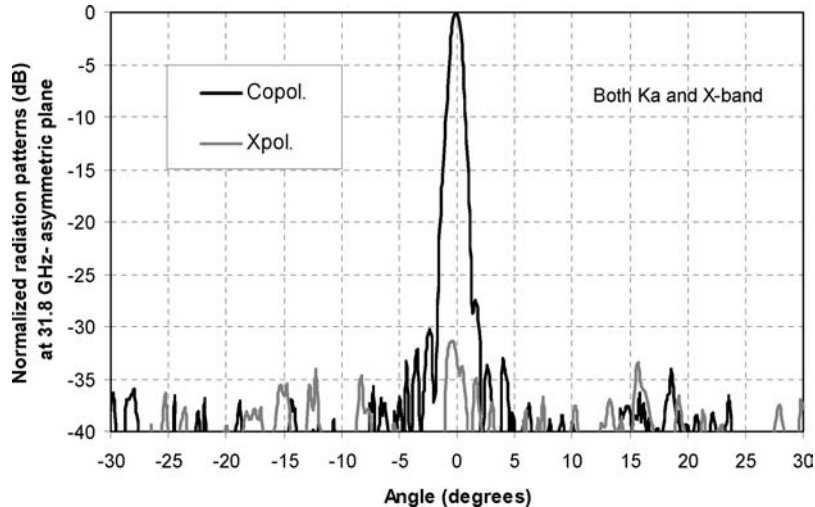


FIGURE 35-23 Measured Ka-band radiation pattern of the X/Ka dual-band reflectarray

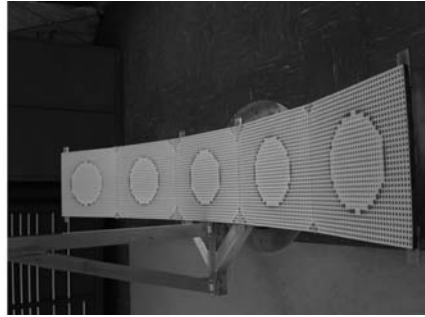


FIGURE 35-24 Piece-wise flat reflectarray for space application

as shown in Figure 35-24, consists of five flat subapertures that are connected together to form a curved reflectarray.³⁵ The curving of the long dimension of the rectangular surface is to minimize the incident angles from the feed for the end elements and, thus, to optimize the radiation efficiency for all elements. The radiation efficiency here indicates the measure of the amount of energy of each element that is reradiated in the desired main beam direction. The advantage of using reflectarray with flat subapertures is to allow mechanical folding of the flat panels into a compact structure for spacecraft launch-vehicle stowage. Test data indicates that this reflectarray is functioning properly with good radiation patterns and an aperture efficiency above 50 percent.

Beam Scanning Reflectarray

One of the key advantages of a reflectarray is that its elements can be implanted with low-loss phase shifters and achieve wide beam scanning without the use of a complicated beam-former (power divider) and expensive T/R amplifier modules. As mentioned in Section 35.2, in the mid 1990s a 94-GHz monolithic reflectarray¹⁷ fabricated in a single wafer, using 1-bit PIN diode phase shifters, achieved wide-angle ($\pm 45^\circ$) electronic beam scanning. Also, a 35-GHz reflectarray, using waveguide/dielectric elements with 3-bit ferrite phase shifters,¹⁸ was reported to achieve $\pm 25^\circ$ beam scanning. More recent developments in this area include ferroelectric thin-film phase shifters^{36,37} for reflectarray. Interest in ferroelectric thin-film circuits developed because of their high power-handling capability, negligible DC power consumption, small size, and potential for low cost. Current research effort is to reduce the phase shifter's relatively high insertion loss (from many decibels to 1 or 2 dB), in particular at the millimeter-wave frequency band.

Another development worth mentioning here is the use of a miniature motor to rotate each element of a circularly polarized reflectarray. This concept was proposed in 1995¹³ and has not been investigated in detail until recently, in which case two miniature motors with two annular-ring reflectarray elements were used in a waveguide simulator³⁸ as well as a 5×1 -element array³⁹ to demonstrate the beam scan.

A third recent important development is the use of the tunable varactor diode^{40,41,42} to control the phase change, as illustrated in Figure 35-25. The phase is analog controlled by its variable control voltage. It not only provides more precision phase but also minimizes the number of control lines when compared to the conventional quantized switched-line phase shifters. For instance, a tunable varactor diode only needs a single variable-voltage control to achieve a full 360° of phase variation, while a 3-bit switched line needs 12 control lines with 45° of quantization error. It is important to minimize the number of control/bias lines in a reflectarray system where generally a huge number of elements is involved.

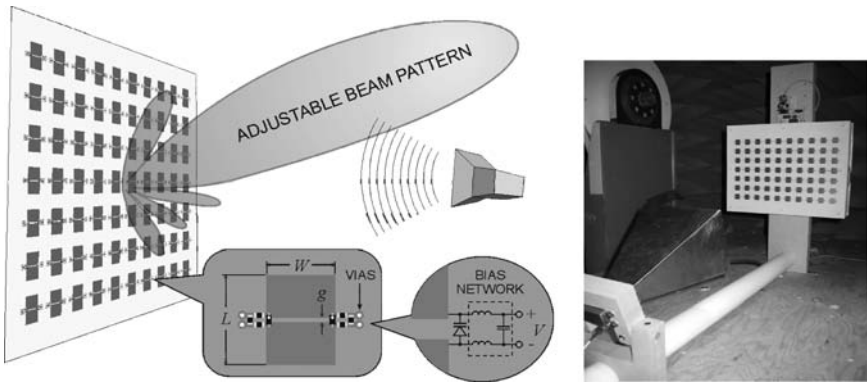


FIGURE 35-25 Beam scanning reflectarray using varactor diodes as phase shifters, with photo showing a C-band 70-element reflectarray using varactor diodes (Courtesy of Sean Hum, University of Calgary, Canada)

35.6 SUMMARY

The reflectarray antenna technology has come a long way. However, its development and application had not been widely adapted until about a decade ago when the printable microstrip reflectarray was introduced. Except for its narrow bandwidth characteristic, the reflectarray has many advantages over a parabolic reflector antenna type. The main beam of a reflectarray can be designed to tilt to a large angle from its broadside direction. Phase shifters can be implanted into the elements for wide-angle electronic beam scanning. For large-aperture spacecraft antenna applications, the reflectarray's flat surface allows the antenna to be made into an inflatable structure with relative ease in maintaining its surface tolerance compared to a curved parabolic surface. Its flat surface also can be made of multiple flat panels for ease in folding into a more compact structure for launch vehicle stowage. Very accurate contour beam shape can be achieved with phase synthesis technique for Earth coverage application. Due to these multitudes of capabilities, the door has just opened for the development, research, and application of the printed reflectarray antennas. Major areas that need continuing improvement of the reflectarray performance are its bandwidth, radiation efficiency, and beam scanning capability.

Acknowledgment

Portions of the research activities described in this chapter were carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

1. D. G. Berry, R. G. Malech, and W. A. Kennedy, "The Reflectarray Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-11 (Nov. 1963): 645–651.
2. H. R. Phelan, "Spiralphase Reflectarray for Multitarget Radar," *Microwave Journal*, vol. 20 (July 1977): 67–73.

3. C. S. Malagisi, "Microstrip Disc Element Reflect. Array," *Electronics and Aerospace Systems Convention* (Sept. 1978): 186–192.
4. J. P. Montgomery, "A Microstrip Reflectarray Antenna Element," *Antenna Applications Symposium*, University of Illinois, Sept. 1978.
5. R. E. Munson and H. Haddad, "Microstrip Reflectarray for Satellite Communication and RCS Enhancement and Reduction" (1987): U.S. Pat. 4,684,952.
6. J. Huang, "Microstrip Reflectarray," *IEEE AP-S/URSI Symposium Digest* (June 1991): 612–615.
7. T. A. Metzler, "Design and Analysis of a Microstrip Reflectarray," Ph.D. dissertation, University of Massachusetts, September 1992.
8. Y. Zhang, K. L. Wu, C. Wu, and J. Litva, "Microstrip Reflectarray: Full-wave Analysis and Design Scheme," *IEEE AP-S/URSI Symposium* (June 1993): 1386–1389.
9. R. D. Javor, X. D. Wu, and K. Chang, "Beam Steering of a Microstrip Flat Reflectarray Antenna," *IEEE AP-S/URSI Symposium* (June 1994): 956–959.
10. D. C. Chang and M. C. Huang, "Multiple Polarization Microstrip Reflectarray Antenna with High Efficiency and Low Cross-Polarization," *IEEE Trans. Antennas Propagat.*, vol. 43 (Aug. 1995): 829–834.
11. A. Kelkar, "FLAPS: Conformal Phased Reflecting Surfaces," *Proc. IEEE National Radar Conf.* (March 1991): 58–62.
12. D. M. Pozar and T. A. Metzler, "Analysis of a Reflectarray Antenna Using Microstrip Patches of Variable Size," *Electronics Letters* (April 1993): 657–658.
13. J. Huang, "Bandwidth Study of Microstrip Reflectarray and a Novel Phased Reflectarray Concept," *IEEE AP-S/URSI Symposium* (June 1995): 582–585.
14. J. Huang and R. J. Pogorzelski, "A Ka-band Microstrip Reflectarray with Elements Having Variable Rotation Angles," *IEEE Trans. Antennas Propagat.*, vol. 46 (May 1998): 650–656.
15. F. S. Johansson, "A New Planar Grating-Reflector Antenna," *IEEE Trans. Antennas Propagat.*, vol. 38 (Sept. 1990): 1491–1495.
16. Y. T. Gao and S. K. Barton, "Phase Correcting Zonal Reflector Incorporating Rings," *IEEE Trans. Antennas Propagat.*, vol. 43 (April 1995): 350–355.
17. J. M. Colin, "Phased Array Radars in France: Present and Future," *IEEE Symposium on Phased Array System and Technology* (Oct. 1996): 458–462.
18. A. A. Tolkachev, V. V. Denisenko, A. V. Shishlov, and A. G. Shubov, "High-Gain Antenna System for Millimeter-Wave Radars with Combined Electrical and Mechanical Beam Steering," *IEEE Symposium on Phased Array System and Technology* (Oct. 1996): 266–271.
19. J. Huang and A. Faria, "A 1-m X-band Inflatable Reflectarray Antenna," *Microwave and Optical Technology Letters*, vol. 20 (Jan. 1999): 97–99.
20. J. Huang, "Capabilities of Printed Reflectarray Antennas," *IEEE Symposium on Phased Array System and Technology* (Oct. 1996): 131–134.
21. D. M. Pozar, S. D. Targonski, and R. Pokuls, "A Shaped-Beam Microstrip Patch Reflectarray," *IEEE Trans. Antennas Propagat.*, vol. 47 (July 1999): 1167–1173.
22. J. A. Encinar, "Design of Two-Layer Printed Reflectarray Using Patches of Variable Size," *IEEE Trans. Antennas Propagat.*, vol. 49 (Oct. 2001): 1403–1410.
23. J. Huang, V. A. Faria, and H. Fang, "Improvement of the Three-Meter Ka-band Inflatable Reflectarray Antenna," *IEEE AP-S/URSI Symposium* (July 2001): 122–125.
24. M. Bialkowski, A. W. Robinson, and H. J. Song, "Design, Development, and Testing of X-band Amplifying Reflectarrays," *IEEE Trans. Antennas Propagat.*, vol. 50 (Aug. 2002): 1065–1076.
25. R. E. Zich, M. Mussetta, M. Tovaglieri, P. Pirinoli, and M. Orefice, "Genetic Optimization of Microstrip Reflectarrays," *IEEE AP-S/URSI Symposium* (June 2002): III-128–131.
26. B. Khayatian and Y. Rahmat-Samii, "Characterizing Reflectarray Antenna Radiation Performance," *IEEE AP-S/URSI Symposium*, Columbus, Ohio, June 2003.
27. J. Huang, C. Han, and K. Chang, "A Cassegrain Offset-Fed Dual-band Reflectarray," *IEEE AP-S/URSI Symposium*, Albuquerque, NM, July 2006.

28. M. Zawadzki and J. Huang, "A Dual-band Reflectarray for X- and Ka-bands," PIRS Symposium, Honolulu, Hawaii, October 2003.
29. W. Menzel, D. Pilz, and M. Al-Tikriti, "Millimeter-Wave Folded Reflector Antennas with High Gain, Low-loss, and Low Profile," *IEEE Antennas & Propagation Magazine*, vol. 44, no. 3 (June 2002): 24–29.
30. D. Pozar, S. D. Targonski, and H. D. Syrigos, "Design of Millimeter Wave Microstrip Reflectarrays," *IEEE Trans. Antennas Propagat.*, vol. 45 (Feb. 1997): 287–296.
31. E. Girard, R. Moulinet, R. Gillard, and H. Legay, "An FDTD Optimization of a Circularly Polarized Reflectarray Unit Cell," *IEEE AP-S/URSI Symposium* (June 2002): III-136–139.
32. J. Huang, "Analysis of a Microstrip Reflectarray Antenna for Microspacecraft Applications," *JPL TDA Progress Report No. 42–120* (Feb. 15, 1995).
33. H. Fang, M. Lou, J. Huang, L. M. Hsia, and G. Kerdanyan, "An Inflatable/Self-Rigidizable Structure for the Reflectarray Antenna," 10th European Electromagnetic Structure Conference, Munich, Germany, Oct. 2001.
34. R. Hodges and M. Zawadzki, "Design of a Large Dual Polarized Ku-band Reflectarray for Spaceborne Radar Altimeter," *IEEE AP-S Symposium* (June 2005): 4356–4359.
35. J. A. Encinar and J. A. Zornoza, "Three-layer Printed Reflectarrays for Contoured Beam Space Applications," *IEEE Trans. Antennas Propagat.*, vol. 52 (May 2004): 1138–1148.
36. R. R. Romanofsky, J. T. Bernard, F. W. Van Keuls, F. A. Miranda, G. Washington, and C. Canedy, "K-band Phased Array Antennas Based on BST Thin-Film Phase Shifters," *IEEE Trans. Microwave Theory and Tech.*, vol. 48 (Dec. 2000): 2504–2510.
37. F. Xiong and R. R. Romanofsky, "Study of Behavior of Digital Modulations for Beam Steerable Reflectarray Antennas," *IEEE Trans. Antennas Propagat.*, vol. 53 (March 2005): 1083–1097.
38. A. E. Martynyuk, J. I. M. Lopez, J. R. Cuevas, and Y. K. Sydoruk, "Wideband Reflective Array Based on Loaded Metal Rings," IEEE MTT-S Microwave Symposium, Long Beach, Calif., June 2005.
39. V. F. Fusco, "Mechanical Beam Scanning Reflectarray," *IEEE Trans. Antennas Propagat.*, vol. 53 (Nov. 2005): 3842–3844.
40. L. Boccia, G. Amendola, and G. Di Massa, "A Microstrip Patch Antenna Oscillator for Reflectarray Applications," *IEEE AP-S/URSI Symposium* (June 2004): 3927–3930.
41. S. V. Hum, M. Okoniewski, and R. J. Davies, "Realizing an Electronically Tunable Reflectarray Using Varactor Diode-tuned Elements," *IEEE Microwave and Wireless Components Letters*, vol. 15 (June 2005): 422–424.
42. M. Riel and J. J. Laurin, "Design of a C-band Reflectarray Element with Full Phase Tuning Range Using Varactor Diodes," IEEE AP-S/URSI Symposium, Washington D. C., July 2005.

Chapter 36

Mobile Handset Antennas

Yiannis C. Vardaxoglou

*Professor of Wireless Communications
Loughborough University, UK*

Jim R. James

*Emeritus Professor, Cranfield University, UK
Consultant Engineer, Visiting Professor, Loughborough
University, UK*

CONTENTS

36.1	IMPACT ON ANTENNA DESIGN	36-2
36.2	CELLULAR HANDSET ANTENNA DESIGN ISSUES	36-2
36.3	HELICAL WIRE ANTENNAS AND VARIANTS	36-5
36.4	EVOLUTION OF THE PIFA AND ITS VARIANTS	36-7
36.5	CERAMIC CHIP AND RESONATOR ANTENNAS	36-11
36.6	SAR MEASUREMENT AND MINIMIZATION	36-14
36.7	PROVISION FOR GPS AND BLUETOOTH.	36-20
36.8	MEASUREMENT OF HANDSET ANTENNAS	36-21
36.9	SATCOM HANDSET ANTENNAS	36-22
36.10	FUTURE TRENDS	36-24
36.11	SYMBOLS	36-25

36.1 IMPACT ON ANTENNA DESIGN

Antenna engineering design has many interesting and surprising features, not least of which is its ability to reinvent itself if the demand arises. Such has been the impact of mobile communications on antenna design, resulting in unbelievably compact and densely packaged dipole-like handset antennas housed entirely within the mobile phone case itself.

The dipole and monopole were some of the first antennas developed and, together with fundamental electromagnetic theory, have been well established over the past century. Yet new versions of these generic radiators continue to be analyzed, simulated, developed, and even researched today as is evident from the research journals worldwide. Quite simply, product demand is the driving force for this seemingly endless creation of different physical realizations of what is essentially an antenna capable of giving a dipole-like radiation pattern.

While the vectorial radiation field of an antenna is uniquely defined by the antenna itself, this is not so for typical engineering radiation patterns specified as the modulus of the radiation field whereby phase information is not retained. Furthermore, this specification is often made even less restrictive by only stating limiting values of sidelobes and other pattern features, between which the measured pattern must lie. Bandwidth specifications are similarly pinned down with only a simple limiting level on the S_{11} response, hence little information is demanded about the shape of the bandpass characteristic itself. The nonuniqueness of such specifications thus leaves plenty of scope to recast the antenna in a very different form.

The dipole antenna family, and in fact most other types of antennas, can be reinvented in a different physical form to satisfy the demands of new equipment products. An outstanding example is the microstrip patch antenna, made possible with printed antenna technology. When compared with conventional monopoles and loop antennas, the printed version can be expected to have a somewhat lower radiation efficiency but it has opened the door to many innovative new products requiring a very low profile dipole-like radiator that is compatible with cost-cutting manufacturing processes.

The previous example not only illustrates how new and very different physical realizations can be developed with similar radiation pattern characteristics but also highlights the need for a system's design approach whereby the performance of individual components is not the main criterion but rather the satisfactory performance of the entire system. The deployment of a system approach has gradually become more evident in mobile handset antenna design over some two decades.

The initial thinking was to mount a conventional antenna on the mobile phone case itself, and early examples were physically large and included balanced fed dipoles, monopoles, and loops. The astronomical demand for mobile phones together with the pull of user preference has resulted in the present-day integral handset antenna. Seldom does one ask of the performance of the antennas in isolation where it is understood that the matching and radiation efficiency data may be less impressive than that of conventional antennas. What matters is how it all functions together as a system.

36.2 CELLULAR HANDSET ANTENNA DESIGN ISSUES

Without doubt, integral handset antenna design is very complicated and presents the designer with perhaps the most formidable antenna design criteria demanded by a product. There are numerous design issues for cellular operation that are summarized next, but the integration as a system further complicates matters, and many, if not most, of the design

issues are interrelated, thus there is a high degree of iteration in the design work. The requirements for a global roaming satcom handset antenna have different constraints and are considered separately in Section 36.9.

Antenna Electrical Size

What constitutes an electrically small antenna is well established and seminal contributions include Wheeler,¹ Chu,² and Collin and Rothschild.³ The radian sphere⁴ of radius a is defined as

$$a = \frac{\lambda}{2\pi} \quad (36-1)$$

where λ is the free-space wavelength. If an antenna's physical structure lies within this sphere, then electrically small antenna characteristics will be evident. The effect of shape and volume of the antenna structure has also been considered by Wheeler.¹

For a lossless antenna, the effects are (i) a high Q factor and hence small bandwidth, (ii) a small radiation resistance resulting in a low radiation efficiency, and (iii) likely greater antenna sensitivity to mechanical and electrical tolerances. When the antenna has intrinsic losses, the bandwidth is less narrow.

It is widely acknowledged⁵ that for diminishing a

$$Q \sim 1 / (ka)^3 \quad (36-2)$$

where $k = 2\pi/\lambda$. From an antenna engineering standpoint, the interest^{6,7} over many years has centered on realizing a small antenna in practice that lies on or exceeds the performance limits represented by the Q formula. This topic has been vigorously revisited recently^{7,8} with the upsurge of activity in small antenna design for both mobile and wireless applications.

For handset antennas, it is the lower bands that enforce antenna electrical size constraints. For instance, at 800 MHz, $a = 5.97$ cm and the dimensions of a typical handset are $1 \times 4 \times 10$ cm, so if the antenna has dimensions of $1 \times 2 \times 4$ cm, it lies within the radian sphere. At 1800 MHz, $a = 2.64$ cm, so the situation is less critical. An unknown factor is, of course, the extent to which a given handset antenna couples to the handset circuit components and battery, etc., because this will in effect increase the electrical size of the antenna.

Ground Plane Electrical Size Effects

When a monopole is mounted on a horizontal ground plane (GP) of finite extent, the dipole-like radiation patterns can be corrupted with a variety of pattern effects depending on the electrical size of the GP. An excellent theoretical illustration with measurements has been given⁹ showing the increase in pattern distortion as a function of reduced GP electrical size. The mobile handset GP is enforced mainly by the metallized assembly consisting of the circuit board and its components, which as mentioned above is typically 4×10 cm. Any plastic outer case will have little effect.

Neglecting hand effects, such a GP could act as an electromagnetic counterpoise to a 10-cm monopole mounted in the same plane as the GP at 750 MHz. Operation at a lower frequency is possible using a monopole version of the sleeve dipole¹⁰ to isolate the handset

case, but the resulting antenna has greater height and is not acceptable today where the antenna element is usually mounted on the GP itself. The GP extent in one direction is only 4 cm and some degree of radiation pattern perturbation is likely to be experienced at the lower mobile bands. Consequently, it is common practice in research papers and manufacturers' data sheets to show measured and/or simulated handset antenna radiation patterns with the antenna embedded on the GP.

There continues to be much interest in creating a stand-alone chip type antenna¹¹ that is immune to the proximity of the GP and its components. Another way of addressing the latter requirements concerns a balanced feed arrangement,¹² but since this concept requires two identical antennas in phase opposition, its application¹³ so far has been confined to the higher mobile bands due to size limitations.

Time Varying Pattern Effects

Cellular handset antennas have to function in a rich multipath propagation environment which allows the use of antennas with polarization properties that are not tightly specified. Hand and head movement, together with the angular movement of the handset, creates further time varying effects which, together with multipath effects, are mainly beyond the designer's control. At the lower mobile bands below 1 GHz, the antenna's constrained electrical size commits it to smoother dipole-like patterns with few nulls and thus time varying effects can be better tolerated than at progressively higher frequencies where additional pattern nulls appear. The situation is further assisted by the self-correcting adaptivity of speech communication.

Increasing Band Coverage

It is generally thought that phone users have an appetite for mobile phones having additional functions that can obtain wider news and sales information and provide more games and a host of other facilities like GPS and wireless links. Financial returns from such a multimedia operation are likely to be very significant but at the same time the cost of the mobile handset must be constrained.

Not surprisingly, handset manufacturers aim to achieve several bands from one central antenna not only to maintain cost levels but also because there is little available space to do otherwise. A multifunction antenna is now commonly the stated requirement, and antenna designers have responded with a variety of configurations that demonstrate design concepts and feasibility. Just how far this concept can be extended remains to be seen, but antennas with three bands have been demonstrated.¹⁴ It is evident that the multiband operation may require some trading of other performance parameters such as radiation efficiency and antenna input match levels. This will limit the number of bands that can eventually be obtained from a single antenna.

Radiation Efficiency

Whatever type of antenna is chosen to embed in the handset, the designer will strive to minimize the power lost to dissipation in the antenna itself. A measure of the losses is the antenna radiation efficiency η_a defined by

$$\eta_a = P_{\text{rad}} / (P_{\text{rad}} + P_{\text{loss}}) \quad (36-3)$$

where the radiated power is P_{rad} and the power dissipated in the antenna is P_{loss} . In practice the antenna is likely to need a matching network, and losses in the latter are expressed as a matching network efficiency η_m where

$$\eta_m = \eta_a Q_m / (Q_m + Q_a) \quad (36-4)$$

where Q_m = matching network Q factor and Q_a = antenna Q factor. The combined efficiency of the antenna and its matching network is then $\eta_a \eta_m$.¹⁵ Further losses are of course incurred when the antenna is embedded in the handset, due to coupling to handset components, and finally when operating with the hand and head present, which account for much of the additional power loss.

36.3 HELICAL WIRE ANTENNAS AND VARIANTS

The remarkable property of wire antennas, or indeed any form of linear metal conductors, is that the resonant frequency is mainly determined by the electrical length of the current path. As such, a wire monopole (or dipole) can be compacted into an electrically smaller size. The antenna bandwidth, however, is more affected and increases for compacted structures occupying a greater volume.

A common example where the wire monopole is partially wound into a coil at some location on the wire¹⁶ is shown in Figure 36-1*a*. For handset applications, a spiral coil of wire alone constitutes the antenna (see Figure 36-1*b*) and for small diameters is known as the normal mode helical antenna (NMHA). The device gives dipole radiation patterns when placed on an electrically large GP but has circularly polarized radiation characteristics when the spiral diameter exceeds about one wavelength. The bandwidth, radiation efficiency, and radiation resistance at resonance are significantly reduced but the latter can be restored by tapping into the spiral, as shown in Figure 36-1*c*, at the expense of a more complicated construction. There are a multitude of variations, and in Figure 36-1*d* a second spiral coil is wound in antiphase directly on top of the first spiral coil, creating a large bandwidth but very poor radiation efficiency. This is referred to as the double wound antenna.¹⁷ Figures 36-1*e* and *f* show zigzag-like versions and other designs that can be configured in fractal form or generated by genetic algorithms. Some performance details for spiral antenna devices are compared to those of a wire monopole in Table 36-1.

A compact manufactured NMHA is illustrated in Figure 36-2 where the conducting path is printed into a groove in the cylindrical plastic former to avoid the increased production cost of winding on a wire.

TABLE 36-1 Performance of Spiral-type Antennas Compared with Monopole Antenna

Antenna Type	Resonance (MHz)	Bandwidth (MHz)	Loss (dB)	Efficiency (%)	Normalized Size	Resistance (Ω)
Monopole	150	~20	~0	~100	1	25
NMHA	142	~1.5	1.04	81	0.22	5
Double wound	150	~9	7.8	16	0.18	21

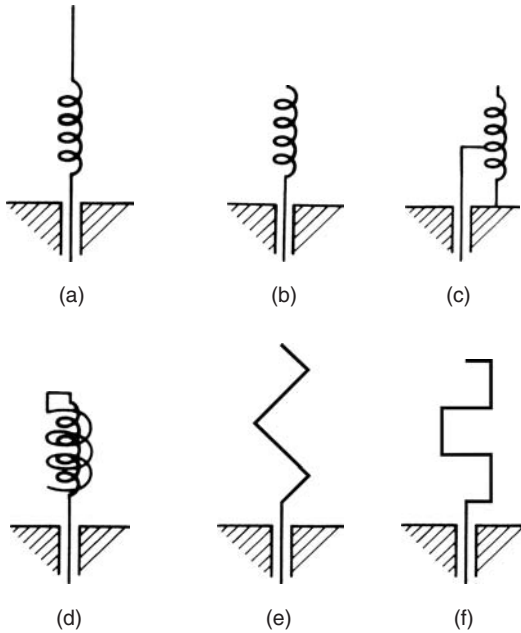


FIGURE 36-1 (a) Coil loaded whip antenna; (b) NMHA; (c) NMHA with tapping connection for matching; (d) double wound spiral antenna; (e) and (f) zigzag antennas. (Figures 16-1a to 1d reproduced by permission of K. Fujimoto and J. R. James¹¹ © Artech House, Inc. 2001.)

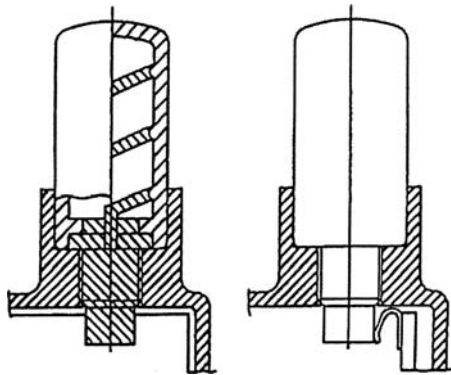


FIGURE 36-2 Example of typical NMHA mounted on a handset case showing construction¹¹ (Courtesy of Nippon Antenna Co. Ltd.)

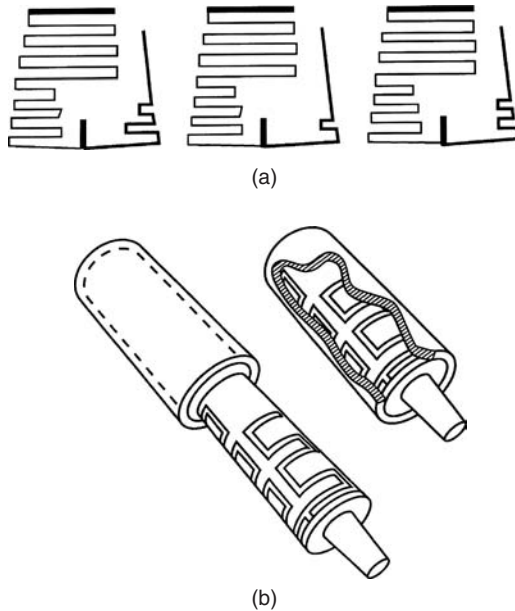


FIGURE 36-3 (a) Printed zigzag antennas printed on dielectric films¹¹ (Courtesy of Allgon); (b) The structure rolled into a cylindrical shape mounted on a plastic former with and without the cover in place

The zigzag antenna configurations have an outstanding manufacturing advantage over the spiral devices because they can be printed on thin dielectric films and then rolled into a cylindrical or other three-dimensional shape. Such assemblies have less size reduction than the spiral types but have attracted use both mounted on a handset case, as shown in Figure 36-3, and embedded within the handset interior as a planar structure. Two printed films designed for different resonant frequencies can also be connected in parallel to obtain a dual-band antenna, and this is more difficult to achieve with spiral conductors.

36.4 EVOLUTION OF THE PIFA AND ITS VARIANTS

The planar inverted F antenna (PIFA) has come into prominence due to its common use in mobile phone handsets. In fact, the PIFA is often regarded as a generic antenna in its own right, but from physical fundamentals it is seen to be a derivative of both wire antennas and printed patch antennas. For instance, a low profile inverted L antenna (ILA) has a low radiation resistance, but by tapping the feed connection along the wire, it can be increased, which is the inverted F antenna (IFA), as shown in Figure 36-4a. This antenna is useful for low profile applications where only a thin wire can be used. Where space is available, the L -shaped wire can be replaced by a planar conductor of width L_1 , as shown in Figure 36-4b. The planar conductor increases the radiation resistance and offers a two-dimensional choice of feed position to facilitate matching. The end shorting plate of height H need not occupy the entire plate width, as shown in Figure 36-4b.

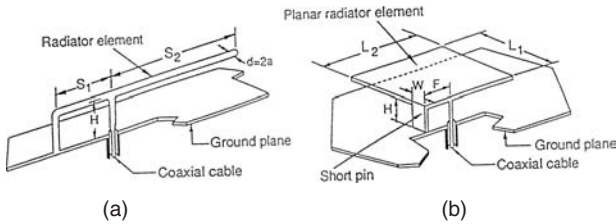


FIGURE 36-4 (a) Inverted F wire antenna and (b) typical PIFA

Examination of Figure 36-4 also reveals that the PIFA can be regarded as a quarter-wave patch antenna¹⁸ with an air-spaced substrate. The PIFA has both vertical and horizontal radiation surfaces, and the radiation patterns can be expected to contain significant cross-polarization, which is not a problem for cellular mobile phones since they have to operate in a high multipath environment. Being purely a simple bent metal antenna with no lossy substrate, the PIFA has good radiation efficiency. However, it is sensitive to its environment, and when embedded in a handset with the presence of the hand and head, the radiation efficiency reduces drastically.¹⁹ The radiation patterns are also significantly perturbed by the hand and head movement and appreciable power is dissipated in the head. However, the simplicity of the PIFA and the low manufacturing cost remain outstanding advantages that are attractive to an industry where cost-cutting is a paramount consideration.

The demand for multiband operation at no additional antenna cost has inspired the creation of the multiband PIFA; a dual-band example is shown in Figure 36-5. The simplicity of this antenna is remarkable and there would appear to be an unlimited number of ways of configuring the current paths on the top plate of the PIFA or indeed to other metal-plate antennas. This concept has been extended to the tri-band version shown in Figure 36-6.

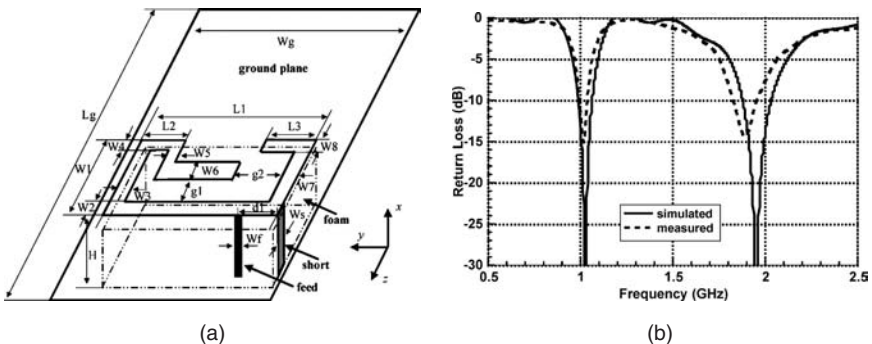


FIGURE 36-5 (a) Dual-band PIFA (after Y.-X. Guo et al²⁰ © IEEE 2004) and (b) bandwidth characteristics of dual-band antenna

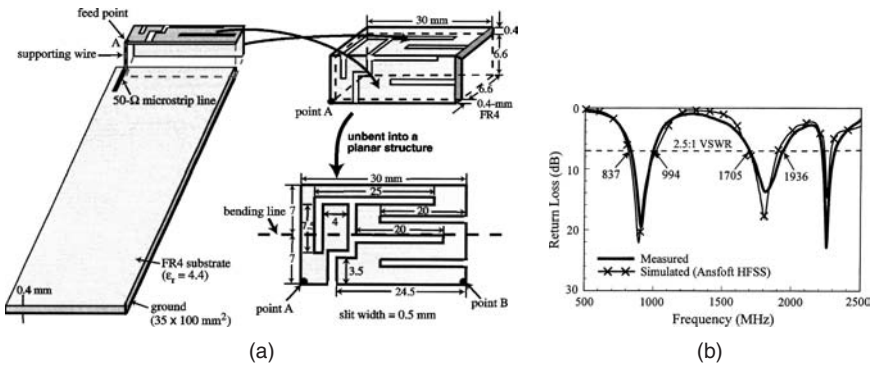


FIGURE 36-6 (a) Tri-band monopole (after G.-Y. Lee and K.-L. Wong²¹ © *Micro. Opt. Technol. Lett.* 2002); (b) Side view of monopole antenna mounted on the handset GP; (c) Measured and simulated bandwidth characteristics of tri-band antenna

As already mentioned, radiation patterns having a high cross-polarization content can be expected, and examples of radiation patterns for the tri-band antenna of Figure 36-6 are shown in Figure 36-7.

Similar performance can be obtained with the meander-line concept, which allows a printed zigzag conductor to be printed on a thin planar dielectric film prior to folding into the rectangular $6 \times 6.5 \times 25$ mm³ box shape illustrated in Figure 36-8. This tri-band antenna is very compact and covers the GSM (900 MHz), DCS (1800 MHz), and PCS (1900 MHz) bands.

The large degree of design freedom offered by these simple metal antenna structures has enabled manufacturers to generate their own in-house low-cost designs. Further compacting of these antennas can be achieved by loading the low-frequency resonant zones of the structure with a high-permittivity dielectric slab, albeit at the expense of a somewhat lower radiation efficiency and bandwidth for the low band and the extra cost of the ceramic material.

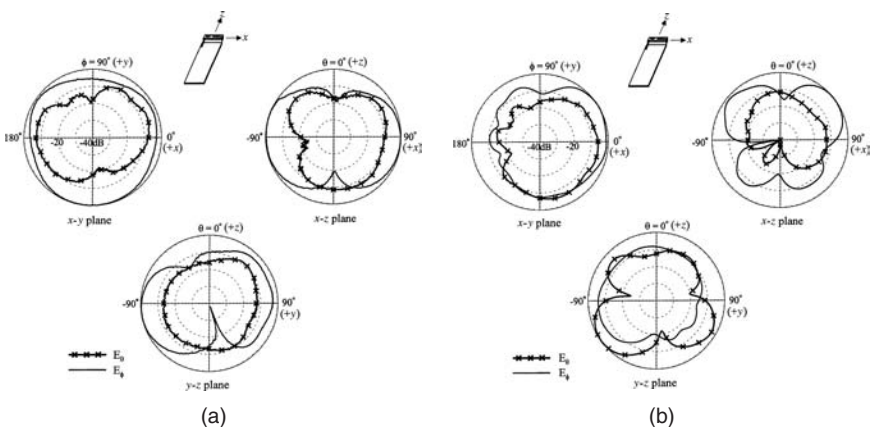


FIGURE 36-7 (a) Radiation patterns of the tri-band antenna at 900 MHz (after G.-Y. Lee and K.-L. Wong²¹ © *Micro. Opt. Technol. Lett.* 2002); (b) Radiation patterns of the tri-band antenna at 1800 MHz

(c)

FIGURE 36-8 (a) The assembled and mounted tri-band meander-line antenna (after P.-L. Teng and K.-L. Wong²² © *Micro. Opt. Technol. Lett.* 2002); (b) The meander-line antenna before assembly; (c) Measured and simulated bandwidth characteristics of the tri-band meander-line antenna

36.5 CERAMIC CHIP AND RESONATOR ANTENNAS

A major manufacturing problem with PIFAs and other air-spaced antennas embedded in the handset is that they are not stand-alone electronic components. These antennas couple to the handset components, and a previously proven antenna cannot be simply embedded into a new handset configuration without further design adjustments. The ceramic chip antenna concept goes some way toward alleviating this coupling problem and also reduces the antenna size at the expense of additional material costs. The fields are intensified within the ceramic material, and unwanted characteristics are the narrowing of the bandwidth together with some lowering of the radiation efficiency even though the intrinsic material loss ($\tan\delta_c$) is low. Numerous chip antenna designs have been described whereby a high-permittivity ceramic material is associated with a meander-line²³ or helical configurations.²⁴ The conducting lines can be readily printed onto the outer surface of the ceramic material. An example of a helical construction is illustrated in Figure 36-9a and its deployment as an antenna embedded in a handset is shown in Figure 36-9b together with radiation patterns.

The typical ceramic chip antenna is likely to be too small to permit resonant modes within its interior. In contrast, the dielectric resonator antenna (DRA)²⁵ relies on a mode being excited, which can be achieved by a small probe or coupled slot. Although the concept of a DRA was introduced by Richtmyer²⁶ in 1939, the practical versions were not investigated until the 1980s, no doubt due to the new availability of high-permittivity, low-loss, temperature-stable ceramic materials at that time. The feasibility of embedding a DRA within a handset has been investigated²⁷ but single mode operation has the characteristic narrow bandwidth which

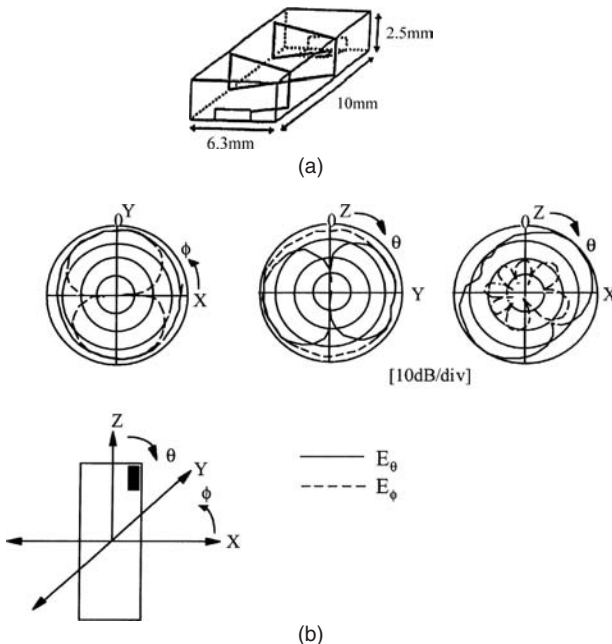


FIGURE 36-9 (a) Constructional view of helical ceramic chip antenna¹¹ (Courtesy of Murata Manufacturing Co.) and (b) the chip antenna embedded in a handset and the radiation patterns

limits its application. In another development²⁸ a quadrifilar-type antenna structure is plated onto a high-permittivity dielectric cylinder. A balun structure at the base isolates the antenna from the GP, and there is a radiation null at broadside, thus reducing radiation exposure to the head. The bandwidth as anticipated is very narrow. The antenna is illustrated in Figure 36-10.

The demand for multiband handset antennas has prompted designers to develop DRAs that have more than one mode excited, thus providing bandwidth windows that correspond to the required cellular bands. This is not difficult if the resonator is not restricted in size, but for handset installation the DRA thickness must be < 10 mm, while the handset width and limited space allocation place severe constraints on the other dimensions. Such a multimode DRA has been described²⁹ and the mode generation is governed by the excitation mechanism, which is commonly a coupling strip plated on the dielectric surface. Additional conductors can be plated on the dielectric surfaces to provide additional bands, and the nature and location of the conductors are mainly determined by experiment and/or simulation.

The electrical compactness of DRAs, at least for the higher mobile bands, allows provision of balanced pairs¹² to reduce currents in the handset GP and hence reduce hand and other environmental effects. An example is given in Figure 36-11, which

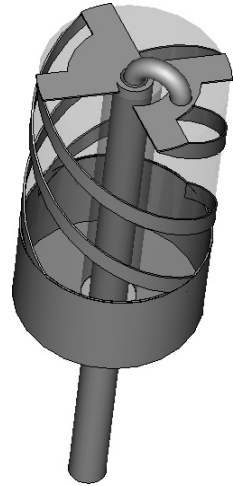


FIGURE 36-10 Dielectrically loaded, balanced quadrifilar-type antenna (Courtesy of Sarantel Ltd.)

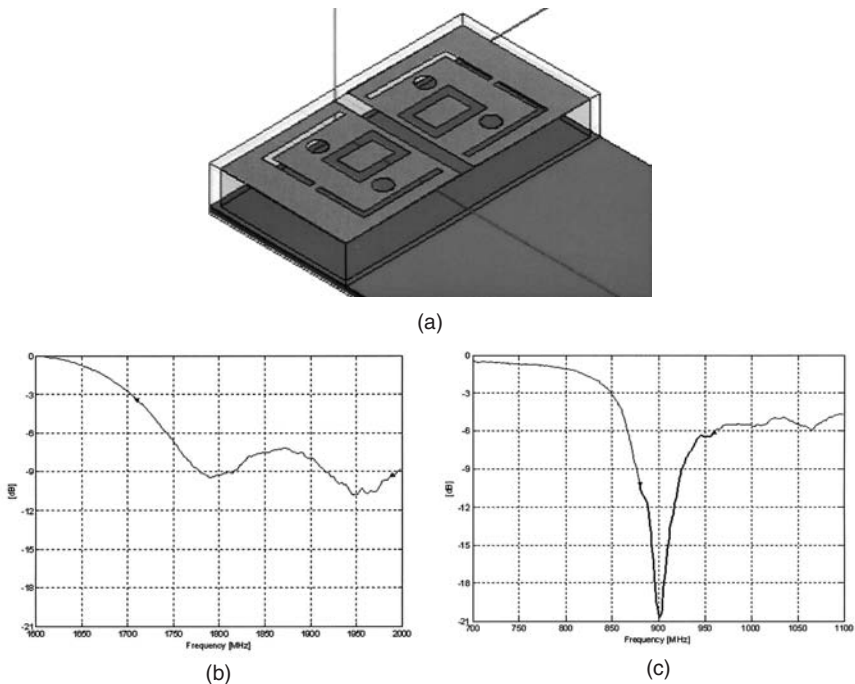
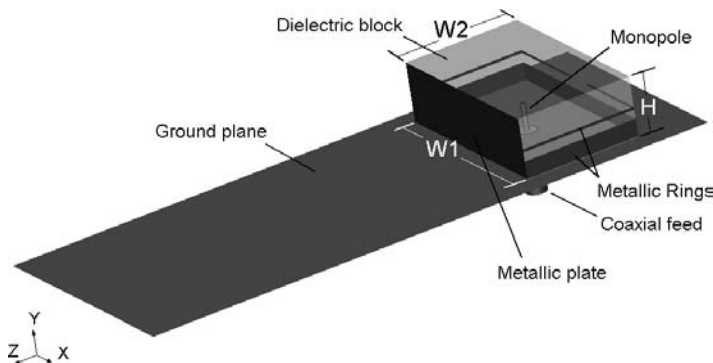


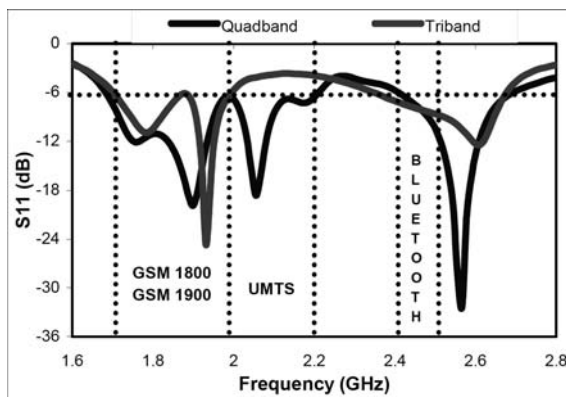
FIGURE 36-11 (a) Balanced ceramic antenna on GP³⁰ (Courtesy of Antenova Ltd.); (b) The 900-MHz bandwidth characteristic; (c) The 1800- and 1900-MHz bandwidth characteristics

covers the 900-, 1800-, and 1900-MHz bands.³⁰ The 900-MHz band remains unbalanced while the two higher bands are balanced. It is also claimed that there is less radiation exposure to the head.

The use of ferrite material to reduce the electrical size of wire monopoles and other antennas has received some attention³¹ over the past 60 or so years but has remained a specialized technique for a few applications. More recently³² its application to DRAs has emphasized the enhancements in antenna performance that can be achieved, these being a significant increase in bandwidth and the radiation efficiency. Best performance is obtained when there is approximate equality between the real parts of the material's complex permittivity and permeability. Theory and simulations³³ have established the performance benefits of applying this technique to handset antennas. An example³⁴ is given in Figure 36-12a, which shows a ferrite-loaded monopole antenna covering the four bands 1800 MHz (GSM), 1900 MHz (GSM), 2100 MHz (UMTS), and 2450 MHz (Bluetooth). The conducting rings facilitate the adjustment of the band shape. The quad-band bandpass characteristics are shown in Figure 36-12b together with a previous tri-band characteristic which did not cover the UMTS band.



(a)



(b)

FIGURE 36-12 (a) View of ferrite antenna on GP (after M. I. Kitra et al³⁴ © IEEE 2007) and (b) the bandwidth characteristics

TABLE 36-2 Performance of Ferrite Handset Antenna (after M. I. Kitra et al³⁴ © IEEE 2007)

	$\tan\delta_\epsilon = 0.03$ $\tan\delta_\mu = 0.03$	$\tan\delta_\epsilon = 0.0001$ $\tan\delta_\mu = 0.06$
ϵ'	6	6.38
μ'	6.6	6.18
Monopole length (mm)	4.5	4.5
Coverage at 1800, 1900, 2100, and 2450 MHz at -6 dB level	Achieved	Achieved
Efficiency (%) with the head present at 1800, 1900, 2100, and 2450 MHz, respectively	35.05, 32.6, 15.15, and 22.34	38.5, 34.9, 19.2, and 22

A summary of the simulated performances for various material properties is given in Table 36-2.

The complex permeability and permittivity of the material are defined as

$$\begin{aligned}\mu_r &= \mu' + j\mu''; & \tan\delta_\mu &= \mu''/\mu' \\ \epsilon_r &= \epsilon' + j\epsilon''; & \tan\delta_\epsilon &= \epsilon''/\epsilon'\end{aligned}\quad (36-5)$$

Simulation using a commercially available ferrite material designed for screening purposes with $\mu' = 2$, $\tan\delta_\mu = 0.06$, $\epsilon' = 24.1$, and $\tan\delta_\epsilon = 0.0001$ established that a quad-band performance was obtainable. The low radiation efficiency may be improved with the development of a ferrite material specifically for antenna rather than screening applications.

36.6 SAR MEASUREMENT AND MINIMIZATION

This section examines the effects on the user's health when using a mobile communications device. Methods for reducing the radiation absorbed in the head are also presented.

Radiation Exposure

The public has been concerned for many years about the electromagnetic radiation effects from microwave ovens, overhead power lines, and household electrical installations even though no established evidence of harm has emerged. Non-ionizing electromagnetic waves in the microwave and lower frequency bands create thermal effects in tissues when the water molecules and dissolved ions are made to vibrate. The water content is therefore an indication of the level of absorption. Nonthermal effects are due to the interaction of the applied field with the molecules where it is considered that the latter align themselves along the electric field to minimize potential energy.³⁵ Elegant theories and mathematical models have been published, for instance, "perturbation of DNA due to soliton waves," "effects on body organs and glands," and "acceleration and retardation of seed growth at specific millimeter wave frequencies." There remains a large amount of literature globally on this topic, but the results are often ambiguous and no clear evidence appears to have emerged so far.³⁶

It is not surprising, therefore, that the astronomical growth of mobile phone usage has been accompanied by much public concern, particularly as the handset is held close to the head and earlier analog handsets could dissipate a watt of power in the brain. Despite extensive worldwide research and strong support from governments for over a decade, no evidence of a health risk has been uncovered so far. Protests about the siting of base station masts and power lines are continuing, but at the present time the situation with mobile phones remains calm and stable, with most users aware that minimizing handset use is desirable, particularly for children. Text messaging has escalated for children and eases the exposure problem.

Manufacturers for their part have reduced exposure levels with the introduction of digital systems and continue to be responsibly mindful of the situation. On the research front, the electromagnetic exposure to the head and other body regions continues to be of major interest, such as the development of low SAR handsets (as described later in this section), the effect of metal spectacles,³⁷ body-worn jewelry,³⁸ metal spectacles on the mucous membranes,³⁹ the exposure to passengers in automobiles when using mobile phones,⁴⁰ to mention but a few. Below we give background details of some current work in this field.

Radio Frequency Dosimetry

With the concern about mobile phone exposure to the human body and particularly the head, it was evident that a regulatory process was urgently needed. The radiation exposure to any part of the human body is now assessed by the Specific Absorption Rate (SAR), which is defined by

$$\text{SAR} = (\sigma / 2\rho) | \mathbf{E} |^2 \text{ W/kg} \quad (36-6)$$

where σ = tissue conductivity, ρ = tissue density, and $| \mathbf{E} |^2$ = the square of the electric field intensity in the tissue. Background details are given in Fujimoto and James¹¹ (Chapter 6, Section 6.5, and Chapter 7). Radio frequency dosimetry in the human body is very complex due to many factors that affect the absorption rate in tissues. Safety guidelines recommend that the 1g averaged peak SAR should not exceed 1.6 W/kg and the whole body averaged peak SAR should be less than 0.08 W/kg. The 1g averaged peak SAR is preferred because it represents local variations more accurately. Each mobile phone user has a usage profile ranging from those who use their mobile phone occasionally to those in a busy occupation necessitating many hours of continuous phone use each day. An idea of how SAR needs to be interpreted is given in Fujimoto and James¹¹ (Chapter 6, Section 6.5). To determine the likely SAR for a mobile phone, it must be measured in as near a natural scenario as possible together with a model of the human head composed of representative artificial tissue. Table 36-3 summarizes the electrical properties of some common tissues.

Purpose-built equipment is available to assist in standardization, but computer modeling has played a major role in rapidly evaluating different human body and handset situations; these aspects are detailed below.

Measurement of SAR

Although computer modeling has proliferated, the direct practical measurement of SAR in a phantom head and body is regarded as an important complementary tool. Handset manufacturers have now built up their bespoke practical measurement equipment over several years with phantom models comprised of tissue simulating fluid with the appropriate electrical parameters as listed in Table 36-3. Elaborate three-dimensional electrically small

TABLE 36-3 Electrical Parameters of Human Body Tissues

Frequency (MHz)	Relative Permittivity ϵ'	Conductivity σ (S/m)
300	45.3	0.87
450	43.5	0.87
835	41.5	0.90
900	41.5	0.97
1450	40.5	1.20
1800	40.0	1.40
1900	40.0	1.40
2000	40.0	1.40
2450	39.2	1.80
3000	38.5	2.40

probes are positioned within the phantom fluid by sophisticated control systems, enabling the SAR to be rapidly assessed within a few minutes. The homogeneous tissue fluid enables a bulk SAR assessment to be performed but cannot represent the layering of different tissues and the presence of organs and glands. It does, however, allow the rapid comparison of different handset antennas and the effects of different handset orientations and separation distance from the body region. Most of the measurements are concerned with the absorption in the head, with or without a phantom hand model, and the variation in SAR for the same handset can be up to 100 percent depending on the measurement conditions.⁴¹ Consumer groups have published league tables of SAR head absorption for commercially available mobile phones but its influence on customer purchasing preference is not very apparent, maybe because of the strong fashion element associated with phones, particularly with younger people.

A photograph of a measurement system is given in Figure 36-13, showing the phantom in the forefront undergoing probe measurements in its right ear region. Details of the probes and control apparatus are given in Fujimoto and James¹¹ (pp. 338–340).

There is also some demand from universities and research establishments for ready-made commercially available SAR measurement equipment. One such example is the SPEAG Dosimetric Assessment System (DASY 4), which provides a six-axes robotic arm scanning the inside of a phantom head with facial features that are filled with tissue simulating liquid.⁴²

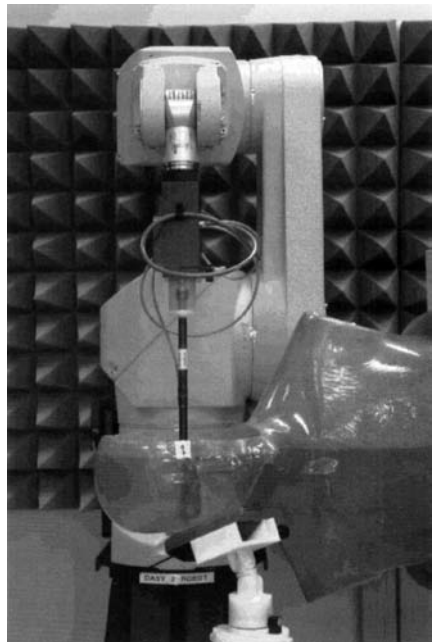


FIGURE 36-13 View of phantom measurement equipment (after K. Fujimoto and J. R. James¹¹ © Artech House, Inc. 2001)



FIGURE 36-14 DASY 4 SAR measurement system by SPEAG

The system is computer controlled, enabling measurement of SAR to be performed in a short time. A photograph of the equipment is shown in Figure 36-14.

Techniques for Minimizing SAR

A variety of earlier attempts sought to reduce the handset antenna radiation to the hand and head regions. A pull-up metal mesh shield around the handset case was manufactured but, as might have been expected, it interfered and reduced the handset transmitted and received power. This idea has very recently been revisited with success using a metamaterial shield⁴³ and a view of the latter in position on the handset case is shown in Figure 36-15.

A way of reducing SAR in the head has been described⁴⁴ that is based on the presence of a ferrite material screen. The research was simulated using FDTD with a detailed head model. It is claimed that provided the handset GP currents are not perturbed, then the SAR in the head is reduced without a reduction in radiation efficiency. An early attempt to reduce the head SAR by creating a radiation pattern null in the direction of the head involved a slot/patch combination.⁴⁵ A layered spherical head constituted the simulation model. The antenna was bulky and unlikely to be applicable to present-day handset requirements, particularly with the demand for single multiband antenna operation.

Computer modeling is now a foremost tool associated with the minimization of SAR, particularly for the head region. The models have ranged from simple homogeneous cube blocks, spheres, and layered spheres to complex, more-realistic head models exhibiting the heterogeneous regions of the head and brain. These anatomical models are very detailed and thus escalate computer processing requirements. Some models can be freely downloaded on the Web. Two different anatomical models^{46,47} are illustrated in Figures 36-16*a* and *b*.

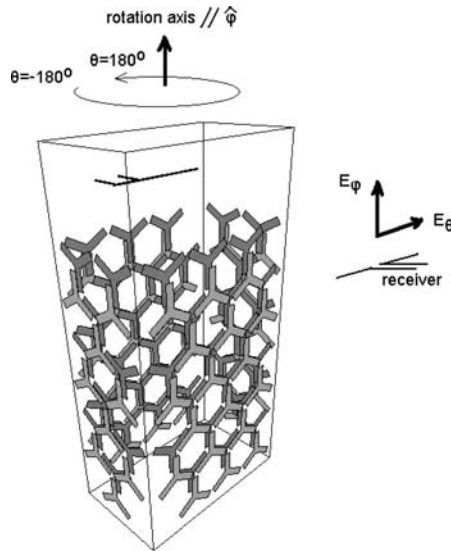
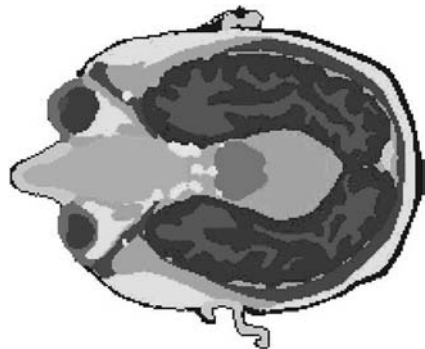


FIGURE 36-15 Metamaterial shield on a handset case (after G.Goussetis et al⁴³ © URSI 2005)

Much research has been carried out on the comparison of different head models and it has been deduced that even though a heterogeneous head model will affect the SAR distribution within the head, it has little effect on the maximum integrated SAR, which is determined mainly by the antenna itself and its distance from the phantom.^{48–50} It is concluded that the maximum averaged SAR is lower in most cases in a heterogeneous head model compared to a homogeneous head model and that a spherical head can be used to give a reasonable, slightly pessimistic assessment.^{50–52}



(a)



(b)

FIGURE 36-16 (a) Visible human head (after U. Tiede et al⁴⁶ © IEEE 1996); (b) 1.1-mm resolution image of human head with 26 tissue types (after W. Whittow⁴⁷ © Sheffield University 2004)

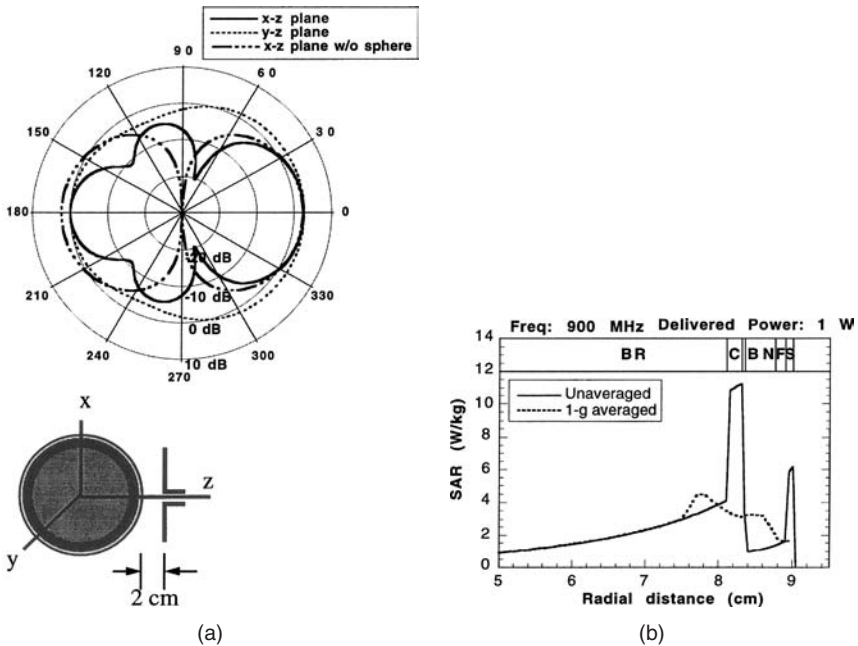


FIGURE 36-17 (a) Far-field gain patterns of the half-wavelength antenna with and without the six-layered spherical head (after K. Fujimoto and J. R. James¹¹ © Artech House, Inc. 2001); (b) 1g averaged and unaveraged SAR distributions along z axis at 900 MHz (BR = brain, C = cerebrospinal fluid [CSF], D = dural, BN = bone, F = fat, and S = skin)

Currently, the IEEE Standards Coordinating Committee 34, Subcommittee 2, Work Group 2 is proposing new FDTD computational techniques for determining the SAR in the human body.⁵³ It is expected that this standard will provide better guidelines for simulations involving heterogeneous heads.⁵⁴

A layered spherical head has been investigated using the eigenvalue expansion method, and an interesting example of the results is shown in Figures 36-17*a* and *b* (Fujimoto and James¹¹ pp. 373–374). The result demonstrates that the absorption in the skin exceeds that in the CSF at higher frequencies.

The majority of head model simulation research appears to mainly involve the use of PIFAs and other handset antennas that couple strongly into their environment, as described earlier in this chapter. The head SAR levels can therefore be near the recommended SAR limits.¹⁹ The ferrite-loaded handset antenna described in Section 36.5 represents a significant advance in techniques for reducing the head SAR, and the reductions in SAR are very significant. Table 36-4 gives details of the SAR in the head for this ferrite antenna. A spherical homogeneous head model was used in this research.

TABLE 36-4 SAR in the Head for Ferrite Handset Antennas³⁵

Handset Antenna	Frequency Band Coverage (MHz)	SAR in the Head (10g [W/kg])
Ferrite antenna (Figure 36-12)	1800, 1900, 2100, and 2450, respectively	0.076, 0.11, 0.06, and 0.0006, respectively

The excellent low SAR in the head is brought about by the use of an antenna monopole excitation probe, the orientation of the probe, and the use of ferrite material, and, for the balanced version, the deployment as well of two coupled antennas in phase opposition. As far as we are aware, these ferrite handset antennas are probably the only way of effectively reducing the SAR in the head while maintaining good prospects for realistic manufacturing potential.

36.7 PROVISION FOR GPS AND BLUETOOTH

The provision for GPS and Bluetooth services in recent years was the commencement of adding value to mobile phones. Handset manufacturers are clearly influenced by the relative ease of embedding additional software processing chips within the handset. Additional bulky sensors are invariably required to input, output, and display the information, and this has not been an insurmountable obstacle for TV reception and camera phones where the display panel and camera can protrude somewhat outside the case. The clamshell handset also eases the space problem. The power of the chips continues to increase significantly while chip prices continue to fall. In keeping with this trend, handset manufacturers are deterred from increasing the number of antennas within the handset, both for cost reasons and because of the acute shortage of space. Therefore, we can expect to have only one multiband handset antenna that will encompass the new bands in addition to the established original bands. The isolation between adjacent bands needs to be small to avoid additional software filtering.

We have already shown in this chapter that some of the multiband antennas are in fact able to include the Bluetooth transceiver band at 2450 MHz. One thing in the designer's favor is that at this higher frequency, the Bluetooth antenna is less electrically small with a quarter-wave length of 3.06 cm. For GPS receiver provision in the L_1 1575.42-MHz band, the antenna size limitation is more critical with a quarter-wave length of 4.76 cm. Both GPS and Bluetooth are low-level signal devices and consequently they involve no SAR issue.

A way of reducing the GPS antenna length has already been demonstrated¹⁹ and involves an IFA plated onto a dielectrically loaded PIFA. The dielectric loading condenses the IFA fields and hence the antenna length. This example is illustrated in Figures 36-18*a* and *b*, showing a view of the dielectric loaded PIFA and the isolation between the PIFA and the IFA.

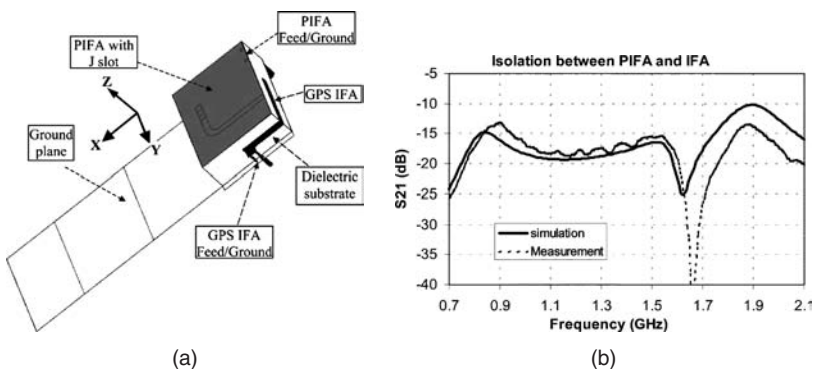


FIGURE 36-18 (a) View of dielectric loaded PIFA with mounted IFA (after Z. Li and Y. Rahmat-Samii¹⁹ © IEEE 2005); (b) Isolation between the PIFA and IFA antennas

How handset manufacturers continue to provide for GPS and Bluetooth depends very much on their own preferred handset and antenna designs, but the examples shown in this section illustrate the feasibility of adding these two services to mobile phones.

36.8 MEASUREMENT OF HANDSET ANTENNAS

The mobile phone handset is a remarkable electronic product: it is packed with communication circuitry that must fit in the handset case, and it must be affordable by the public at large. The complexity of handsets has demanded new thinking about not only how to measure the circuit functions during the manufacturing process but also its overall performance when assembled. All of this must be accomplished in a cost-effective way. The situation is further complicated by the continually increasing processing speeds and the associated high-frequency EMC difficulties within the handset. Instrument manufacturers have responded in earnest and numerous suites of sophisticated instruments are now available to handset manufacturers to automate their component and subsystem production lines.

Measuring the performance of the finished product is a somewhat less-well-defined task because it has to embrace the relationship of the handset with its environment. This includes not only radio propagation effects in the everyday urban scenario but, as already discussed in this chapter, the absorption of waves by the operator's body and, in particular, head. To what extent these issues are addressed will depend on the handset manufacturers but the cost of testing will be a major consideration and it has to be assumed that some sampling of the manufacturing output will take place.

The performance of an antenna in cellular operation can be defined and related to a particular route. Multipath reflections from moving vehicles, people, airplanes, doors in buildings, etc., create random fluctuations in the propagating radio wave polarization and signal strength. For measurement purposes, the wave polarization can be resolved into vertical polarization (VP) and horizontal polarization (HP), so upon moving the mobile handset along a particular urban route, the mean powers P_v and P_h , respectively, associated with these polarized components can be measured. The antenna will not be able to capture all the polarized power, and $P_r < (P_v + P_h)$, where P_r is the mean antenna power that would have been received over the same route, in the same time period. The antenna performance is defined⁵⁵ by its mean effective gain (MEG), where

$$\text{MEG} = P_r / (P_v + P_h) \quad (36-7)$$

$$\text{MEG} = \int_0^{2\pi} \int_0^\pi [P_v / (P_v + P_h) G_\theta(\theta, \phi) P_\theta(\theta, \phi) + P_h / (P_v + P_h) G_\phi(\theta, \phi) P_\phi(\theta, \phi)] \cdot \sin \theta d\theta d\phi \quad (36-8)$$

where (θ, ϕ) are spherical angular coordinates, $G_\theta(\theta, \phi)$ and $G_\phi(\theta, \phi)$ are the θ and ϕ components of the antenna power gain pattern, respectively, and $P_\theta(\theta, \phi)$ and $P_\phi(\theta, \phi)$ are the θ and ϕ components of the angular density functions of the incoming plane waves, respectively.

The recent trend toward extreme handset compactness and internal integrated antennas does not allow the handset antenna to be evaluated in isolation from the handset itself. In principle the MEG assessment could be applied to the handset itself, but it is clearly not affordable as a production test. Measurements on handset reception and transmission are also carried out⁵⁶ in a controlled environment, such as an anechoic chamber, but the process can again be very time-consuming and indeed costly.

The use of a reverberation chamber to test antennas in Rayleigh fading propagation conditions is well established, but its recent application⁵⁷ to mobile handset testing appears to offer manufacturers an efficient, low-cost method of assessing handset performance. An anechoic chamber requires highly absorbent walls, but a reverberation chamber is a large, metal-walled cavity capable of supporting many modes, which are then greatly perturbed by rotating reflectors within the excited chamber. This mode-stirring action is random enough to create a Rayleigh distributed transfer function between the receiving and transmitting antennas inside the chamber: in this application, these are the radiating handset and the receiving wall probe. The Rayleigh distribution model is a useful approximation of realistic mobile propagation conditions, and a demonstration chamber has been described⁵⁷ that has many attractive features. For instance, a phantom tissue head model can be placed in close proximity to the handset under measurement to enable the actual operating radiation efficiency to be obtained. The chamber can also test the addition of diversity functions and the newly emerging MIMO systems. The installation costs and operating times are low and compatible with handset manufacturing requirements.

36.9 SATCOM HANDSET ANTENNAS

Unlike cellular phones, satcom phone systems have had a very uncertain start due to the inadequate public take-up, which has led to financial difficulties for manufacturers. The systems have, however, survived in a leaner, more responsive form and a stable market has been developed. Around 1998 the systems under development were Iridium, Globalstar, ICO, and Thuraya. A summary of their salient technical details are listed in *Mobile Antenna Systems Handbook*¹¹ (p. 582). Iridium, Globalstar, and ICO are based on LEO, LEO, and MEO constellations, respectively, while Thuraya has a GEO satellite. Phased-array antennas directing spot beams to earth are used on all systems. Global roaming is provided by satcom phones, enabling the user to communicate from any place on earth. The user community includes military, business people, politicians, emergency services, those working in remote parts of the world or at sea, and so on, and this service differs greatly from that of the cellular phone system. A point in favor of the handset manufacturers is that the handset cases are less restricted in size than cellular phones, cost is not such a limitation, and the handset antenna can protrude well above the handset case. It was evident⁵⁸ in 1998 that the perceived market may not be adequate although the global roaming capability would attract support from the defense sector.

These satcom systems introduced new problems for handset antenna designers because of the low signal margins and the need for circular polarization.⁵⁸ The quadrifilar helix antenna (QHA) is extensively used, and reduced-diameter versions have been developed, giving some loss of bandwidth capability while ensuring compact user convenience. QHAs are fitted with a balun to feed the spiral arms from an unbalanced feeder, and various designs exist. A more conventional QHA is shown in Figure 36-19.

The current Iridium 9505A handset has a size of $158 \times 62 \times 59$ mm and has many added-value features including fax and data; an auxiliary handset antenna is also supplied. The Globalstar SAT600

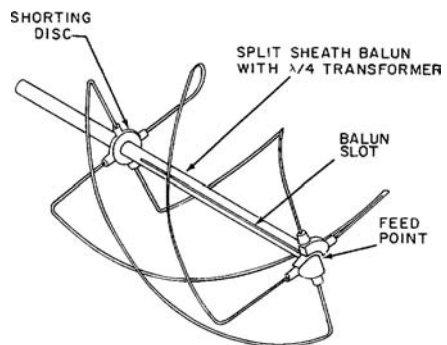


FIGURE 36-19 Quadrifilar helix antenna (QHA) (after C. Kilgus⁵⁹ © IEEE 1964)



FIGURE 36-20 Thuraya 7101 handset¹¹ (Courtesy of Thuraya)

handset has a size of $177 \times 58 \times 39$ mm and offers a dual-mode facility combining cellular GSM900 and the satcom channel. Add-on facilities for the Internet are also available. The Thuraya 7101 handset, shown in Figure 36-20, offers data, fax, and GPS facilities; its size is $146 \times 53 \times 27$ mm.

As already mentioned, the QHA is constructed so that it can be pulled up out of the handset's case for use, thus giving it an elevated position clear of the user's head. Construction details for the Thuraya handset antenna are given in Figure 36-21, which shows the narrow-diameter QHA with the GSM NMHA mounted on its tip.

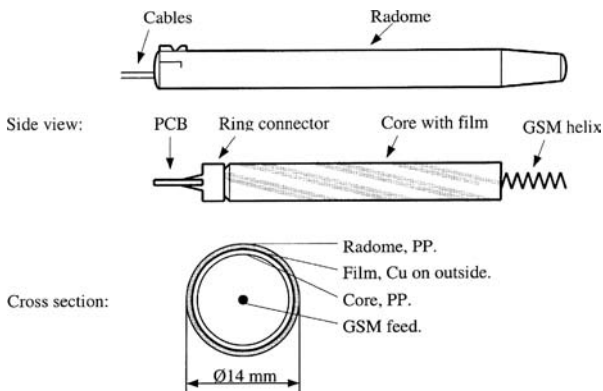


FIGURE 36-21 Constructional details of the Thuraya handset antenna¹¹ (Courtesy of Allgon)

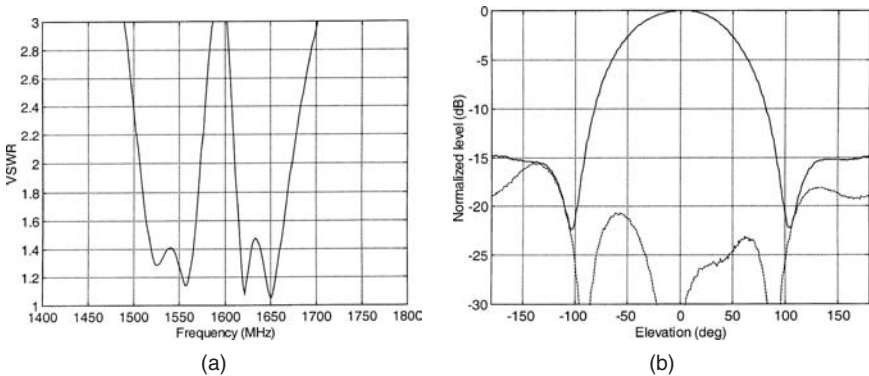


FIGURE 36-22 (a) Thuraya handset antenna passband characteristics¹¹ (Courtesy of Allgon); (b) Thuraya handset antenna's LHCP and RHCP elevation radiation pattern

The Thuraya earth-to-space link is 1626.5–1660.5 MHz with a space-to-earth link of 1525.0–1559.0 MHz. The Thuraya passband characteristics are shown in Figure 36-22a, and the antenna elevation radiation pattern is shown in Figure 36-22b.

36.10 FUTURE TRENDS

The mobile communications market will continue to grow fueled by user demand for increased functionality, aesthetics, smaller handsets, longer battery life, and ubiquitous access. Some future trends are presented next.

- In the immediate future, handset manufacturers will continue to add value to handsets by way of low-cost processing software. This of course will be dependent on the consumer takeup, which from past experience is dictated by individual requirements, fashion, and quite likely other serendipity factors. When the new services require extended band coverage, handset antenna innovation will continue to be much in demand.
- The system design approach will continue to be much in evidence in handset antenna design. This takes a holistic view of how all the components function together even if their individual performance may be degraded. When compared with conventional antennas, present-day handset antennas, with their low efficiency and proximity to other components, would have been seen as very poor radiators a few decades ago.
- Increasing the data rates of mobile phones will continue to receive research attention as more services are sought. New processing algorithms and techniques are being addressed, including the use of multiple antennas and quantum computing.⁶⁰
- The clamshell type of mobile handset has provided more space on handsets for facilities and displays while allowing the handset to be compacted when not in use, but additional antenna design problems are invoked.⁶¹ Other types of handset case designs can be expected as services increase.
- The quest for a single multiband handset antenna that will cover all the band requirements for new services is inhibited by well-established fundamental antenna action whereby antenna resonances are necessary to achieve good matching, bandwidth, and radiation efficiency. Resonance is determined by the length of current paths and sets constraints

on antenna size even if material loading and other compacting techniques are used, as described in this chapter. A future breakthrough in antenna design cannot be ruled out. Much attention has been focused recently on how materials can be synthesized using metamaterials⁶² but to date the cell sizes of the latter remain too large for the synthesis of an electrically small antenna.

36.11 SYMBOLS

A	radius of minimum sphere
DRA	dielectric resonator antenna
GP	ground plane
GPS	global positioning system
IFA	inverted F antenna
K	free-space wave number
LEO	low earth orbit
MEG	mean effective gain
MEO	medium earth orbit
NMHA	normal mode helical antenna
η_a	antenna radiation efficiency
η_m	matching network radiation efficiency
PIFA	planar inverted F antenna
Q	Q factor
Q_a	antenna Q factor
Q_m	matching network Q
QHA	quadrifilar helix antenna
SAR	specific absorption rate
UMTS	universal mobile telephone system
$\tan\delta_\mu$	permeability loss tangent
$\tan\delta_\epsilon$	permittivity loss tangent
$\mu_r = \mu' + j\mu''$	complex permeability
$\epsilon_r = \epsilon' + j\epsilon''$	complex permittivity

REFERENCES

1. H. A. Wheeler, "Fundamental Limitations of Small Antennas," *IRE Proc.*, vol. 35 (December 1947): 1479–1484.
2. L. J. Chu, "Physical Limitations of Omni-directional Antennas," *J. App. Phys.*, vol. 19 (December 1948): 1163–1175.

3. R. E. Collin and S. Rothschild, "Evaluation of Antenna Q," *IEEE Trans. Antennas and Propagat.*, vol. AP 12 (January 1964): 23–27.
4. H. A. Wheeler, "The Radian Sphere Around a Small Antenna," *IEE Proc.*, vol. 47 (August 1959): 1325–1331.
5. J. S. McLean, "A Re-examination of the Fundamental Limits on the Radiation Q of Electrically Small Antennas," *IEEE Trans. Antennas and Propagat.*, vol. 44, no. 5 (May 1996): 672–676.
6. G. Goubau, "Multielement Monopole Antennas," *Proc. ECOM_ARO Workshop on Electrically Small Antennas* (May 6–7, 1976): 63–67.
7. R. C. Hansen, "Fundamental Limitations in Antennas," *Proc. IEEE*, vol. 69 (February 1981): 170–182.
8. A. D. Yaghjian and S. R. Best, "Impedance, Bandwidth and Q of Antennas," *IEEE Trans. Antennas and Propagat.*, vol. 53, no. 4 (April 2005): 1298–1324.
9. D. M. Bolle and M. D. Morganstern, "Monopole and Conic Antennas on Spherical Vehicles," *IEEE Trans. Antenna. and Propagat.*, vol. 17 (1969): 477–484.
10. A. W. Rudge, K. Milne, A. D. Olver, and P. Knight (eds.), *The Handbook of Antenna Design*, vol. 2 (London: Peter Peregrinus, IEE, 1983): 727.
11. K. Fujimoto and J. R. James, *Mobile Antenna Systems Handbook*, 2nd Ed. (Boston: Artech House 2001).
12. H. Morishita and K. Fujimoto, "A Balanced-Fed Loop Antenna System for Handset," *IEICE Trans. Commun.*, vol. E82-A, no. 7 (1999): 1138–1143.
13. S. Kingsley, "Advances in Handset Antenna Design," http://rfdesign.com/mag/radio_advances_handset_antenna/ (May 1, 2005): 16–22.
14. K. Wong, *Planar Antennas for Wireless Communications* (Hoboken, NJ: Wiley-InterScience, 2003): 53.
15. K. Fujimoto, A. Henderson, K. Hirasawa, and J. R. James, *Small Antennas* (Baldock, UK: Research Studies Press, 1987): 7–9.
16. R. C. Hansen, "Optimum Inductive Loading of Short Whip Antennas," *IEEE Trans.*, vol. VT-24 (1975): 21–29.
17. J. R. James and A. Henderson, "Investigation of Electrically Small VHF and HF Cavity-type Antennas," *Proc. Int. Conf. Ant. Propagat.* (1978): 322–326.
18. J. R. James and P. S. Hall, *Handbook of Microstrip Antennas*, vols. 1 and 2 (London: Peter Peregrinus, IEE, 1989): 25, 1104.
19. Z. Li and Y. Rahmat-Samii, "Optimization of PIFA-IFA Combination in Handset Antenna Design," *IEEE Trans. Antennas and Propagat.*, vol. 53, no. 5 (May 2005): 1770–1778.
20. Y.-X. Guo, M. Y. W. Chia, and Z. N. Chen, "Miniature Built-In Multiband Antennas for Mobile Handsets," *IEEE Trans. Antenna. and Propagat.*, vol. 52, no. 8 (August 2004): 1936–1944.
21. G.-Y. Lee and K.-L. Wong, "Very-Low-Profile Bent Planar Monopole Antenna for GSM/DCS Dual-band Mobile Phone," *Microwave and Optical Technol. Lett.*, vol. 34, issue 6 (September 2002): 406–409.
22. P.-L. Teng and K.-L. Wong, "Planar Monopole Folded into a Compact Structure for Very-Low-Profile Multiband Mobile-Phone Antenna," *Microwave and Optical Technol. Lett.*, vol. 33, issue 1 (April 2002): 22–25.
23. J. Lee, C. Jeon, and B. Lee, "Design of Ceramic Chip Antenna for Bluetooth Applications Using Meander-Lines," *IEEE Antennas and Propagat. Symposium* (2002): 68–71.
24. C. Lin, Y. Cheng, and H. Chuang, "Design of a 900/1800 MHz Dual-Band LTCC Chip Antenna for Mobile Communication Applications," *Microwave Journal* (January 2004): 78, 80, 82, 84, 86.
25. K. M. Luk and K. W. Leung, *Dielectric Resonator Antennas* (Baldock, UK: Research Studies Press Ltd, 2003).
26. R. D. Richtmyer, "Dielectric Resonators," *Jour. Appl. Phys.*, vol. 10 (June 1939): 391–398.
27. M. T. K. Tam and R. D. Murch, "Compact Sector and Annular Dielectric Resonator Antennas," *Trans. Antennas and Propagat.*, vol. 47, no. 5 (May 1999): 837–842.

28. O. Leisten et al., "Miniaturised Dielectrically Loaded Quadrifilar Antenna for Global Positioning System (GPS)," *Electronics Letters*, vol. 37, issue 22 (October 2001): 1321–1322.
29. G. Ba-Babik, C. Di Nallo, and A. Farone, "Multi-mode Dielectric Resonator Antenna of Very High Permittivity," *IEEE Int. Conf. Antennas and Propagat.* (2004): 1383–1386.
30. Antenova Ltd., GSM RADIONOVA Radio Antenna Module Data Sheet, February 4, 2005.
31. J. R. James and A. Henderson, "Electrically Short Monopole Antennas with Dielectric or Ferrite Coatings," *Proc. IEE*, vol. 125, no. 2 (1978): 793–803.
32. J. R. James, R. Chair, K. M. Luk, K. M. Chow, K. W. Leung, and J. C. Vardaxoglou, "Influence of Magnetic Material on Dielectric Resonator Antenna Excitation," *IEE Proc. Microw. Antennas and Propagat.*, vol. 151, no. 4 (August 2004): 293–298.
33. M. I. Kitra, P. McEvoy, J. C. Vardaxoglou, and J. R. James, "A Theoretical and Simulation Study of Dielectrically Loaded Antennas and Their Contribution Towards Low SAR," *Inter. ITG Conf. on Antennas (INICA)* (September 2003): 245–248.
34. M. I. Kitra, C. J. Panagamuwa, P. McEvoy, J. C. Vardaxoglou, and J. R. James, "Low SAR Ferrite Handset Antenna Design," *IEEE Trans. Antenna and Propagation*, vol. 55, no. 4 (April 2007): 1155–1164.
35. A. de Salles, "Biological Effects of Microwave and RF," *Proc. SBMO/IEEE MMT-S IMOC '99*, (April 1999): 51–56.
36. R. Goldberg, "Literature Resources for Understanding Biological Effects of Electromagnetic Fields," *EMF-link Multimedia Resource*, <http://infoventures.com/emf/top/lit-rev.html> (January 1996).
37. W. G. Whittow and R. M. Edwards, "A Study of Changes to Specific Absorption Rates in the Human Eye Close to Perfectly Conducting Spectacles Within the Radio Frequency Range 1.5 to 3.0 GHz," *IEEE Trans. Antennas and Propagat.*, vol. 52, no. 12 (2004): 3207–3212.
38. N. A. Samsuri and J. A. Flint, "On the Effect of Jewellery Rings on Specific Absorption Rate (SAR)," *Proc. Loughborough Antennas and Propagat. Conf.* (2006): 421–423.
39. W. G. Whittow and R. M. Edwards, "A Study of Changes to Specific Absorption Rates in the Mucous Membrane Close to Perfectly Conducting Spectacles Within the Radio Frequency Range 0.8 to 2.6 GHz" *Proc. Loughborough Antennas and Propagat. Conf.* (2006): 417–420.
40. D. O. Coy, D. M. Zakharia, and Q. Balzano, "Field Strengths and Specific Absorption Rates in Automotive Environments," *IEEE Trans. Vehicular Tech.*, vol. 48, no. 4 (1999): 1287–1303.
41. J. C. Lin, "Specific Absorption Rates (SARs) Induced in Head Tissues by Microwave Radiation from Cell Phones," *IEEE Antennas and Propagat. Mag.*, vol. 42, no. 5 (October 2000): 138–139.
42. *DASY 4 Dosimetric Assessment System Manual*, Schmidt and Partner Engineering AG, 2003.
43. G. Goussetis, A. P. Feresidis, G. K. Palikaras, M. Kitra, and J. C. Vardaxoglou, "Miniaturisation of Electromagnetic Band Gap Structures for Mobile Applications," *Radio Science*, vol. 40, no. 6, RS6S04 (Nov. 2005).
44. J. Wang, O. Fujjwar, and T. Takagi, "Reduction of Electromagnetic Absorption in the Human Head for Portable Telephones by a Ferrite Sheet Attachment," *IEICE Trans. Commun.*, vol. E80-B, no. 12 (1997): 1810–1815.
45. H. Ruoss and F. M. Landstorfer, "Slot Antenna for Handheld Mobile Phones Showing Significantly Reduced Interaction with the Human Body," *Electron. Lett.*, vol. 32, no. 6 (1996): 513–514.
46. U. Tiede, T. Schiemann, and K. H. Höhne, "Visualizing the Visible Human," *IEEE Comput. Graphics Appl.*, vol. 16 (1996): 7–9.
47. W. G. Whittow, "Specific Absorption Rate Perturbations in the Eyes and Head by Metallic Spectacles at Personal Radio Communication Frequencies," Ph.D. thesis, EEE Dept., University of Sheffield, UK, 2004.
48. S. Watanabe, M. Taki, T. Nojima, and O. Fujiwara, "Characteristics of the SAR Distributions in a Head Exposed to EM Fields Radiated by a Hand-held Portable Radio," *IEEE Trans. on Microw. Theory and Tech.*, vol. 44, no. 10 (October 1996): 1874–1883.
49. A. Lee, H. Choi, B. Kim, H. Lee, and J. Pack, "Effect of Head Size for Mobile Phone Exposure on EM Absorption," *Proc. APMC2001* (2001): 384–387.

50. K. Kim and Yahya Rahmat-Samii, "Antennas and Human in Personal Communications: Applications of Modern EM Computational Techniques," *Proc. 12th Inter. Conf. on Microw. and Radar*, vol. 4 (May 20–22, 1998): 36–55.
51. S. Khalatbanri, D. Sardaari, A. A. Mirzaee, and H. A. Sadafi, "Calculating SAR in Two Models of the Human Head Exposed to Mobile Phone Radiation at 900 and 1800 MHz," *Prog. in Electromag. Res. Symp.*, Cambridge, Mass., March 26–29, 2006.
52. K. Kim and Yahya Rahmat-Samii, "EM Interactions Between Handheld Antennas and Humans: Anatomical vs Multi-layered Spherical Head," *IEEE Trans. APS Conf. on Antennas and Propagation for Wireless Communications* (November 1998): 69–72.
53. "Recommended Practice for Determining the Peak Spatial-Average Specific Absorption Rate (SAR) in the Human Body from Wireless Communications Devices, 30 MHz–6 GHz: General Requirements for Using the Finite Difference Time Domain (FDTD) Method for SAR Calculations," IEEE 1528.1 draft standard, April 2006.
54. B. B. Beard et al, "Comparisons of Computed Mobile Phone Induced SAR in the SAM Phantom to That in Anatomically Correct Models of the Human Head," *IEEE Trans. Electro. Comp.*, vol. 48, no. 2 (May 2006): 397–407.
55. T. Taga, "Analysis for Mean Effective Gain of Mobile Antennas in Land Mobile Antenna Environments," *IEEE Trans. Veh. Tech.*, vol. VT-39, no. 2 (May 1990): 117–131.
56. H. Arai, *Measurement of Mobile Antenna Systems* (Boston: Artech House, 2001).
57. P. S. Kildal, "Characterisation of Small Antennas and Active Mobile Terminals in Rayleigh Fading by Using Reverberation Chamber," *Proc. Loughborough Antennas and Propagat. Conf.* (April 2005): 234–239.
58. J. R. James. "Realising Personal Satcom Antennas," *IEE Electronics and Communications Eng. Jour.* (April 1998): 73–82.
59. C. C. Kilgus, "Spacecraft and Ground Station Applications of the Resonant Quadrifilar Helix," *IEEE Ap-S Int. Symp. Digest* (June 1964): 75–77.
60. R. Calderback, "Quantum Computing and Cellular Phones," Lecture at The Royal Academy of Engineering, London, June 20, 2006.
61. B. S. Collins, "Improving the Performance of Clamshell Handsets," *Proc. Loughborough Antennas and Propagat. Conf.* (2006): 7–12.
62. R. W. Ziolkowski and A. D. Kipple, "Application of Double Negative Materials to Increase the Power Radiated by Electrically Small Antennas," *IEEE Trans. Antenna. and Propagat.*, vol. 51, no. 10 (October 2003): 2626–2640.

Chapter 37

**Broadband Planar
Antennas for
High-Speed Wireless
Communications**

Zhi Ning Chen

Institute for Infocomm Research, Singapore

CONTENTS

37.1	INTRODUCTION	37-2
37.2	SUSPENDED PLATE ANTENNAS	37-2
37.3	PLANAR MONOPOLE ANTENNAS	37-11

37.1 INTRODUCTION

High data speeds and multiband operations of modern wireless communication systems have significantly increased the demand for broadband antennas that are capable of supporting such requirements. In general, the systems need low-cost solutions with desired performance in terms of impedance bandwidth, polarization, and gain. Owing to unique merits such as small volume or low profile, low manufacturing cost, and easy integration into planar circuits, planar antennas are playing important roles in various wireless communication applications.

The planar antennas can usually be categorized in terms of radiation performance into microstrip patch antenna, suspended plate antenna (SPA), planar inverted-L/F antenna, and planar monopole/dipole antenna.¹ Usually, they are attractive to antenna engineers due to their low profile and/or small volume. The changes in such antenna design are, in general, from the specific requirements of applications. For example, the microstrip patch antenna in its basic forms has a low profile, which is conducive to conformal design, but suffers narrow impedance bandwidth on order of 1 percent. In contrast, the planar monopoles usually have a high profile above a ground plane but enjoy broad bandwidth.

Making a trade-off between antenna profile and impedance, as well as radiation performance, the SPAs are good options for fixed base stations in wireless communication systems, and planar monopoles/dipoles for mobile wireless terminals. Accordingly, a variety of techniques have long been developed to further enhance the broadband performance of the SPAs and planar monopoles. Due to the merits of acceptable performance, low profile, and, in particular, low manufacturing cost, the SPAs and planar monopoles have widely been applied in high-speed wireless communication systems.

This chapter is divided into two parts, suspended plate antennas, and planar monopoles for high-speed wireless communications. The basic characteristics of the antennas will be introduced. The state-of-the art designs will be covered, and the important issues for the antenna applications will be addressed.

37.2 SUSPENDED PLATE ANTENNAS

In this section, the SPAs will be introduced as base station antennas for modern broadband wireless connections such as high-speed wireless local area networks (WLANs) based on IEEE 802.11. The impedance and radiation characteristics of the SPAs are described first. Then, the techniques used to broaden the impedance bandwidth of the SPAs are addressed. After that, the techniques for enhancing the radiation performance of the SPAs are introduced. Finally, the applications of the SPAs to WLANs will be exemplified as a case study.

Definition

An SPA with a thick, low-permittivity dielectric substrate can be considered as a variation of microstrip patch antenna, as shown in Figure 37-1.¹ The planar radiator may be of any shape and placed above a ground plane. The planar radiator may be supported by a dielectric substrate with very low permittivity. Besides air, the dielectric substrate may be low-loss dielectric foam with a relative dielectric constant, ϵ_r , of ~ 1 . The dielectric substrate may be multilayered partially with air. The substrate thickness h is usually around 0.06 times the operating wavelength. The antenna may be fed in various ways. In Figure 37-1

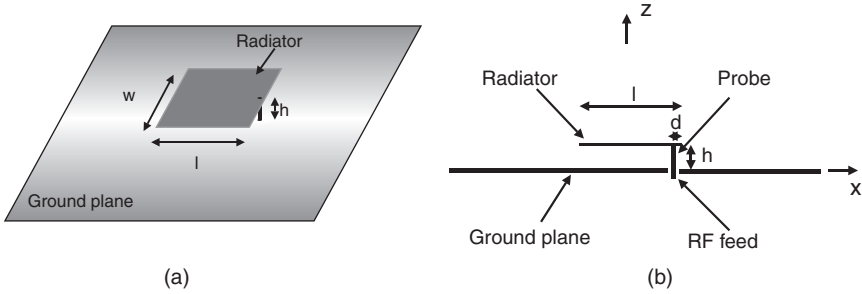


FIGURE 37-1 Coordinate system and geometry of a probe-fed SPA above air: (a) three-dimensional view and (b) side view

the SPA is fed by a coaxial probe at a distance d away from the edge of the radiator. The operating frequency f_o of the SPA can be estimated by

$$f_o = \frac{c}{\lambda_o} = \frac{c}{2\sqrt{\epsilon_r}(l+h)} \quad (37-1)$$

where λ_o denotes the operating wavelength at the frequency f_o , c is the speed of light (299,792,458 m/s), and ϵ_r designates the relative dielectric constant. The length l , width w , distance d , and thickness h are in meters.

The SPAs can be fed in many ways like microstrip patch antennas. Typically, the SPAs can be excited by a microstrip line, a probe, and an aperture-coupled feeding structure, as shown in Figure 37-2.

Figure 37-2a describes the very popular feeding structure of the planar antennas, where a stripline is directly connected to the edge of the SPA.² Different from the microstrip patch antenna with a thin dielectric substrate supporting the radiator, the stripline for the SPAs is usually suspended above a ground plane with a large thickness. The large spacing between the feeding stripline and ground plane causes a wide stripline of characteristic impedance of 50 Ω . For example, within a frequency range of 1–10 GHz, the width of a stripline is 5.1 mm for $h = 1$ mm and 10 mm for $h = 2$ mm. Therefore, a coaxial probe, namely the inner conductor of a coaxial connector, is used to vertically (not horizontally) excite the end of the stripline. As a result, the frequency-dependent input impedance of the probe is observed.

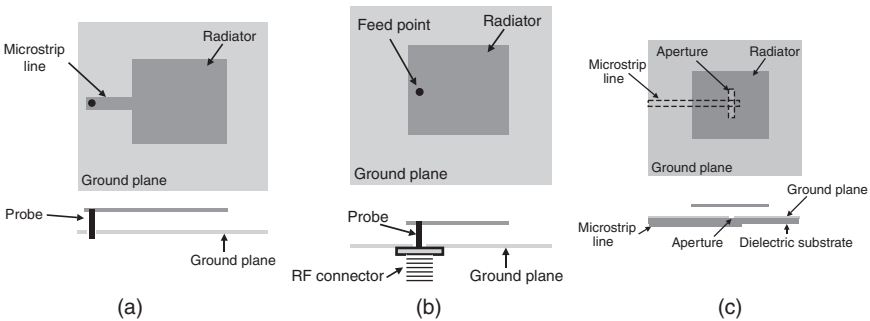


FIGURE 37-2 Three types of feeding structures for SPAs

In Figure 37-2*b*, the SPA is fed by a coaxial probe attached to the radiator directly while its outer conductor is connected to the ground plane.³ This configuration greatly simplifies the feeding structure. Due to the large spacing between the radiator and ground plane, this is a proper solution for the SPAs. The feeding probe can also be used to support the radiator when no dielectric substrate is used below the radiator. However, both the stripline and coaxial probe feeds may cause narrow impedance bandwidth due to the narrowband impedance transition between the stripline or probe and radiators. Moreover, such asymmetric feeding structures degrade the radiation performance, especially at higher frequencies with high cross-polarization radiation levels and/or squinted radiation from the boresight axis (beam-squint effect).

Figure 37-2*c* shows an electromagnetically coupled feeding structure,⁴ where a microstrip line centrally feeds the radiator through a nonresonant slot right below the radiator, which is cut from the ground plane. The symmetric configuration of the antenna to some degree alleviates the degraded radiation performance problems by suppressing the high cross-polarization radiation and the severe beam-squint effect. The nonattaching transition structure also improves the impedance bandwidth to some extent.

Important Features

Consider the SPA and coordinate system shown in Figure 37-1 again. The radiator is a square with a length $l = 58$ mm. A coaxial probe feeds the radiator with a distance d away from the radiator edge. Figure 37-3 shows the input impedance of the probe when the thickness h is 5, 8, and 11 mm. The start and stop frequencies are 2 and 3 GHz, respectively. From the Smith chart in Figure 37-3, the variation of the input impedance for varying thickness h is clearly observed. When the thickness becomes larger, the input impedance becomes larger, in particular, with larger inductance. This is an important feature of the SPAs. Therefore, to achieve a broad impedance bandwidth, it is necessary to employ the techniques to compensate for the large inductance.

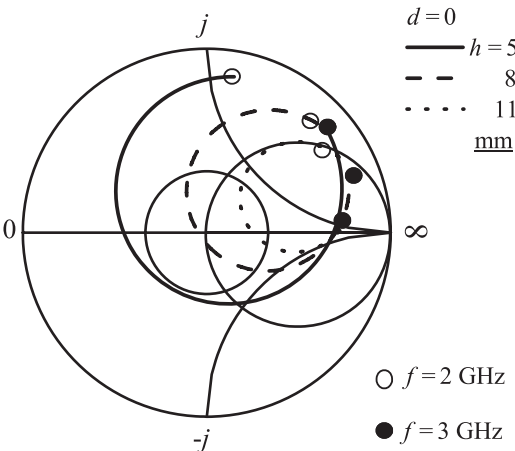


FIGURE 37-3 Input impedance of the probe-fed SPA shown in Figure 37-1 with $d = 0$ mm, $l = 58$ mm, and $h = 5, 8$, and 11 mm

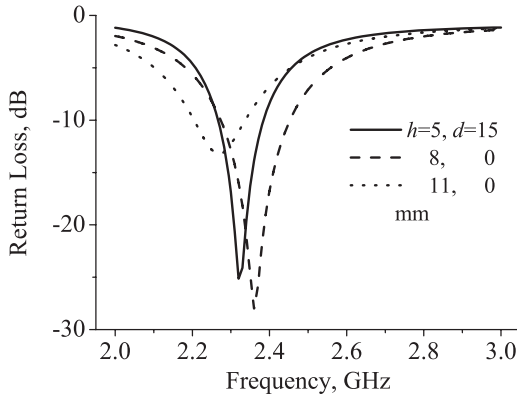


FIGURE 37-4 Return losses of the SPA shown in Figure 37-1 with $l = 58$ mm and $h = 5, 8$, and 11 mm.

Figure 37-4 shows the return losses of the SPA shown in Figure 37-1. The location of the feed point is optimized for good impedance matching as $h = 5$ mm. The achieved impedance bandwidths for -10 -dB return loss are, respectively, 4.7 percent, 7.1 percent, and 6.2 percent at $h = 5, 8$, and 11 mm. The larger the thickness, the broader the impedance bandwidth. However, for the larger thickness, such as 11 mm, the good matching cannot be achieved by changing the position of the feedpoint, so that the achieved impedance bandwidth is not larger as expected. Therefore, the impedance-matching techniques should be applied for broad impedance bandwidth.

Owing to the large thickness, the radiation performance of the SPAs is slightly different from that of the conventional microstrip patch antennas. Figure 37-5 shows the radiation patterns in both E and H planes at the lower-edge, center, and higher-edge frequencies of the impedance bandwidth. In the E planes, the cross-polarization radiation levels are very low and thus can be ignored. With the increase in operating frequency, the sidelobe levels greatly increase, and the beam-squint effect can be observed at the higher-edge frequency due to the asymmetric configuration of the antenna. In the H planes, the copolarized radiation patterns are hardly changed. However, the cross-polarization radiation levels increase by 5.4 dB when increasing the frequency.

In short, the longer probe results in higher impedance, in particular, inductance. The larger antenna thickness offers potential for a broad impedance bandwidth but degrades the radiation performance with higher cross-polarization radiation levels in the H planes and higher sidelobe levels as well as the beam squinting in the E planes. As a result, it is necessary to develop the techniques to broaden impedance bandwidth and enhance the radiation performance in the applications of the SPAs in wireless communication systems.

Techniques for Broad Impedance Bandwidth

In the SPA design, the large thickness and low permittivity of the substrate provides the potential for a broad impedance bandwidth due to low antenna Q values. For example, the probe-fed SPAs in their basic forms with air as their substrate usually can achieve around 7 percent impedance bandwidth for -10 -dB return loss.¹ The achieved bandwidth is limited by the high impedance caused by the longer feeding probe. To further broaden the impedance bandwidth, the techniques for impedance matching have been developed and applied.¹

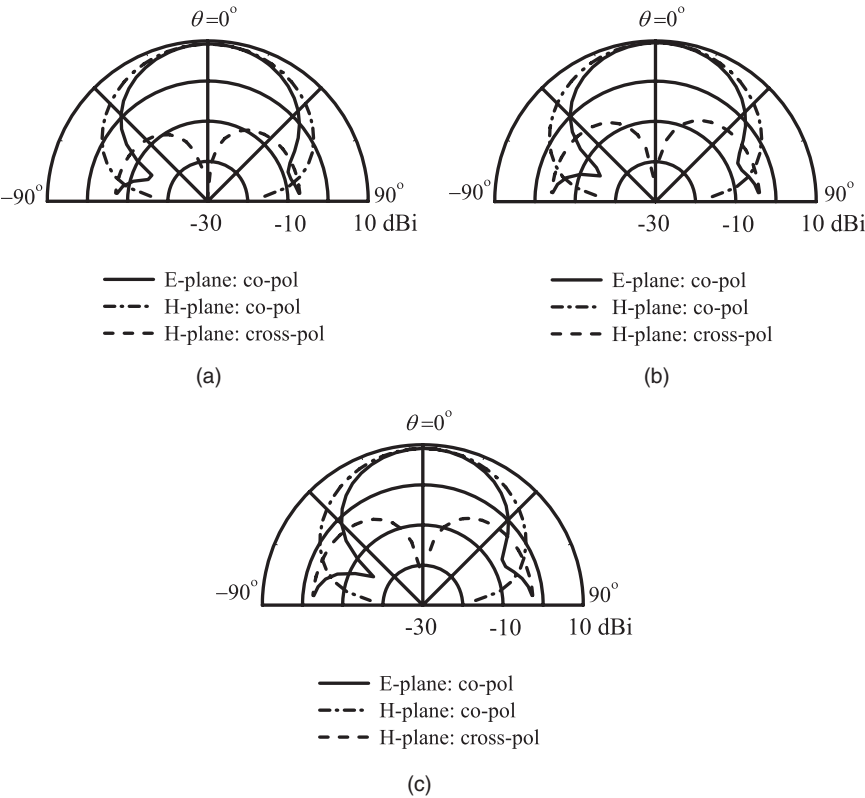


FIGURE 37-5 Radiation patterns for the probe-fed SPA shown in Figure 37-1 with $l = 58$ mm and $h = 8$ mm at (a) lower-edge frequency, 2.28 GHz; (b) center frequency, 2.36 GHz; (c) higher-edge frequency, 2.44 GHz

Figure 37-6 shows the techniques for broadening impedance bandwidths of the probe-fed SPAs, which are extensively used in wireless communication systems. Figure 37-6a shows the method to offset the large inductance caused by a longer probe.⁵ An annular slot is cut around the feedpoint so that the radiator is fed through capacitive coupling. The capacitive loading can also be introduced between the radiator and feeding probes in other ways.⁶⁻⁸

Figure 37-6b illustrates the technique to counterbalance the large inductance and achieve a broad bandwidth by slotting the radiator.⁹ The slot cut from the radiator acts as an impedance-matching network across a broad bandwidth. The slot can be of other shapes.¹⁰ Such a technique can be applied without any increase in the size of antenna and can easily be fabricated at a low cost; two factors that make the applications of planar antenna arrays appealing.

The introduction of strong electromagnetic coupling between the feeding probe and radiator is an important and effective way to achieve broadband impedance matching.¹¹ The idea has been applied to the SPAs with great success, as shown in Figure 37-6c. Due to the application of the L-shaped feeding probe, the impedance bandwidth can reach up to more than 36 percent.¹²

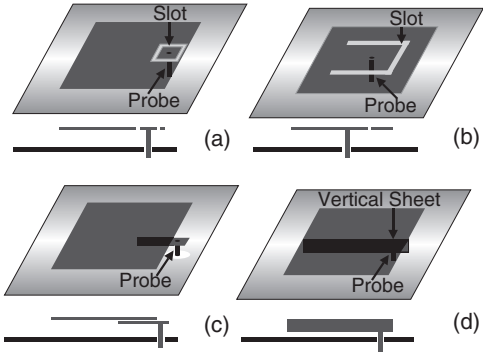


FIGURE 37-6 Techniques for broadening impedance bandwidth of the SPAs

The key to achieving the broadband impedance response of the SPA is the broadband transition between the feeding probe and the radiator. Figure 37-6d shows the other way to modify the transition, where a vertical conducting sheet is used between the feeding probe and the radiator.¹³ The vertical sheet basically acts as a broadband impedance-matching network.

As an example, an SPA fed by an L-shaped probe is designed that operates at 2.4 GHz. The geometry of the SPA is shown in Figure 37-7a, where a rectangular stub (25×4 mm) forms the horizontal portion of the L-shaped probe and is vertically excited by a probe with a radius of 0.65 mm. By adjusting the length and height of the vertical portion of the L-shaped probe as well as the height of the radiator, the impedance locus forms a tight loop around the center of the Smith chart, as shown in Figure 37-7b. The strong electromagnetically coupled structure functions as a broadband impedance-matching network. Therefore, the well-matched bandwidth for –10-dB return loss reaches up to 25 percent, ranging from 2.04 GHz to 2.62 GHz.

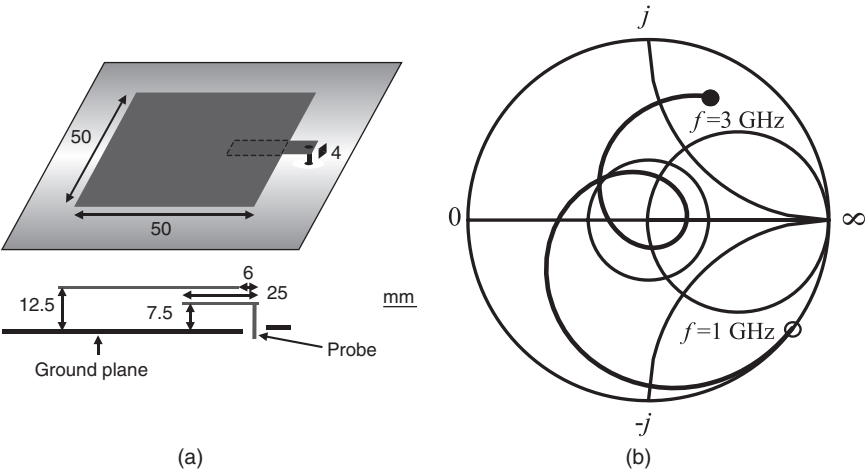


FIGURE 37-7 (a) Geometry of an L-shaped probe-fed SPA and (b) impedance in Smith chart

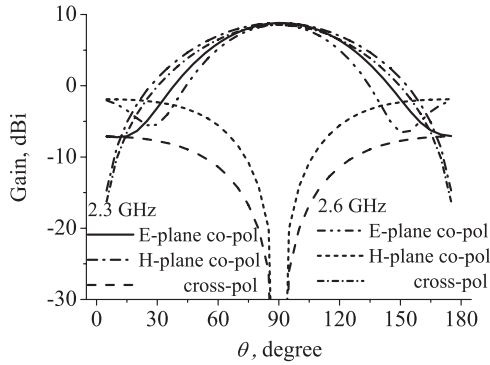


FIGURE 37-8 Radiation patterns for gain of the L-shaped probe-fed SPA.

Figure 37-8 shows the radiation patterns in both E and H planes at frequencies 2.3 and 2.6 GHz. With the increase in frequency, the copolarization sidelobe levels in the E planes increase by 5.2 dB. The beam-squint effect is observed at the higher frequency of 2.6 GHz. In the H planes, the cross-polarization radiation levels become high also by 5.2 dB at 2.6 GHz. The ratios of co- to cross-polarization radiation levels are reduced from 15.8 dB at 2.3 GHz to 10.4 dB at 2.6 GHz. Therefore, the L-shaped probe-fed SPA features a broad impedance bandwidth but suffers the degraded radiation performance.

Techniques for Enhanced Radiation Performance

For base station antennas in wireless communication systems, the pure radiation polarization and stable maximum radiation direction are often required to avoid possible interference between different wireless systems and to save radiation power. The techniques to enhance the radiation performance of the broadband SPAs have been developed. As mentioned in *Broadband Planar Antennas: Design and Applications*,¹ the structurally asymmetric and electrically unbalanced antenna design usually suffers high cross-polarization radiation levels and beam-squint effects. Consequently, the solution to alleviate the degraded radiation performance problem is to modify the feeding structure of the SPAs to keep the antenna geometry symmetric and feeding structure balanced.

Figure 37-9 demonstrates some of the solutions that can be easily applied to the SPA design with the enhanced radiation performance. First, a dual-probe feeding configuration is illustrated in Figure 37-9a.¹⁴ Two feeding probes are used to excite the symmetric radiator simultaneously. The probes are symmetrically located in the E plane with respect to the H plane so that the antenna configuration is kept geometrically symmetric. The out-of-phase RF signals from the two probes form an electrically balanced feeding structure. As a result, the currents that distribute symmetrically and in balance at the radiator effectively suppress the cross-polarization radiation levels in the H planes and beam-squint effect in the E planes.

However, the dual-probe feeding scheme needs a broadband feeding network to form a feeding system with two anti-phase probes, although this can be eased in array design.¹⁵ To simplify the dual-probe feeding structure, a single-probe scheme is presented in Figure 37-9b.¹⁶ The radiator is fed by a probe through a half-wavelength strip. The probe

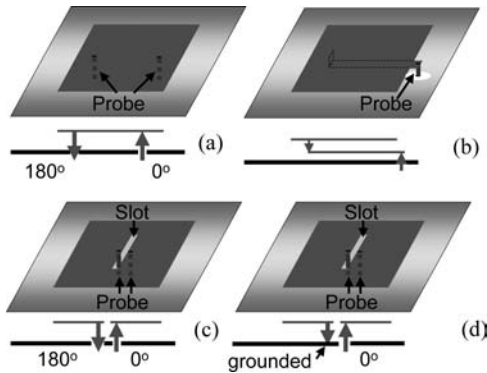


FIGURE 37-9 Feeding schemes applied to SPAs for enhancement of radiation performance: (a) dual-probe; (b) probe-fed half-wavelength strip; (c) dual-probe-fed center slot (d) simplified dual-probe-fed center slot scheme.

vertically excites one of the ends of the strip. The horizontal portion of feeding strip is half-wavelength long at the center operating frequency. Another end of the strip is vertically attached to the planar radiator. The RF signals in the two ends are expected to be out of phase. There is strong electromagnetic coupling between the horizontal portion of the strip and radiator, which achieves a good impedance matching within a broad bandwidth of 20 percent. Due to the balance-like feeding configuration, the cross-polarization radiation levels are effectively reduced so that the ratio of co- to cross-polarization levels is higher than 20 dB across the whole bandwidth.

The beam-squint effect in the E planes is, however, observed due to the asymmetric geometry caused by the feeding structure. A modified dual-probe feeding structure is used to enhance the radiation performance, as shown in Figure 37-9c, where the two out-of-phase probes are positioned around the center of the radiator, and a narrow slot is used to separate two feeding points.^{17,18} The separation between the two probes is small. Varying the length of the center slot can shift the resonant frequency and improve impedance matching. As discussed in *Broadband Planar Antennas: Design and Applications*,¹ the current distribution at the radiator will be kept symmetric so that the cross-polarization radiation is effectively suppressed in all directions. Such a structure also needs a feeding network in the single element design similar to the conventional dual-probe scheme. A modified version was proposed and used as shown in Figure 37-9d, where a grounded pin is used to replace the second probe. The radiation performance of the SPAs with such a simplified feeding scheme is comparable with that shown in Figure 37-9c.¹⁹

Applications in High-Speed Wireless Communications

As mentioned in previous sections, the SPAs feature broad impedance bandwidth and acceptable radiation performance. One more important merit is that the SPAs have a cost-effective design, because they can be readily fabricated with conducting sheets partially or completely without use of any expensive dielectric substrate. Such an advantage enhances the competence of such

an antenna technology in the massive wireless market. The application of the SPA in wireless communication systems, for example, the access point (AP) of WLANs, is exemplified with an actual product design.

As is well known by now, wireless networks can offer many advantages over wired networks due to mobility. With wireless networks, users are free from the tether of an Ethernet cable at a desk. Also, the wireless networks are flexible, fast, and easy to use, deploy, and maintain. The WLAN is a mature option for wireless networks, which are based on the IEEE 802.11 protocol and widely used. One of the WLAN's operating frequency bands (IEEE 802.11b) is the 2.4-GHz band (2.4–2.4835 GHz for North America and Europe; 2.471–2.497 GHz for Japan; 2.4465–2.4835 GHz for France; and 2.445–2.475 GHz for Spain). Thus, the antennas for the WLANs should cover the band of 2.4–2.5 GHz as well.

Figure 37-10a shows a photo of one design that has been successfully applied in the APs of WLANs as a product. The SPA was formed by a piece of copper sheet and supported by two stands above a 100×100-mm ground plane at a distance of 7.5 mm, as shown in Figure 37-10b. The stands were symmetrically cut and bent from the radiator and separated by a center slot. One stand was grounded as a shorting pin and the other was fed by a 50-Ω microstrip line through the ground plane. Adjusting the length of the slot can change the operating frequency and improve impedance matching. The detailed optimized parameters are shown in Figure 37-10b. The antenna is embedded in a casing of the APs for indoor WLAN wireless access applications.

Figure 37-11 shows the measured return loss and radiation patterns for the SPA. The bandwidth for -15 dB can cover the operating band of 2.4–2.5 GHz very well, as shown in Figure 37-11a. The measured radiation patterns in both the E and H planes at 2.4 and 2.5 GHz in Figure 37-11b and c show the gain in the boresight direction is higher than 8.5 dBi. The wider coverage of 76° in the H planes and slight tilt beam direction of $5\text{--}10^\circ$ in the E planes are conducive to application of the SPA design in the indoor WLAN AP applications, where the APs are operating at the mode of point-to-multiple point.

It is concluded that the suspended plate antenna is one type of broadband microstrip antenna but with many unique features such as larger antenna thickness, broader impedance bandwidth, more easily degraded radiation performance, and lower fabrication cost. A variety of the techniques for the enhancement of impedance and radiation performance have long been developed. The more cost-effective designs based on suspended plate antennas have been accepted by the massive wireless communication markets.

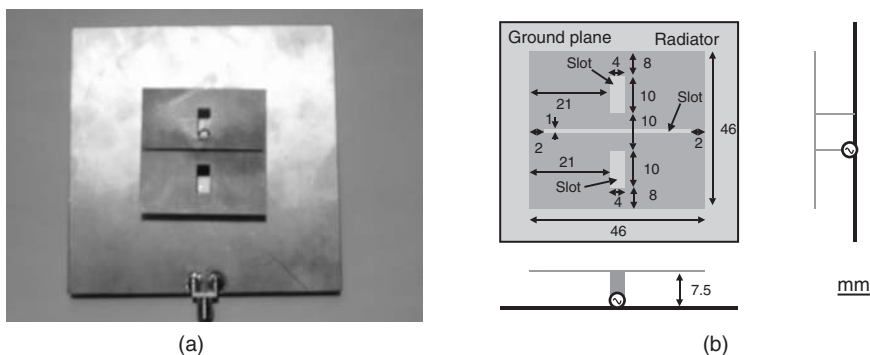


FIGURE 37-10 Geometry of an SPA embedded in a WLAN access point operating at 2.4 GHz

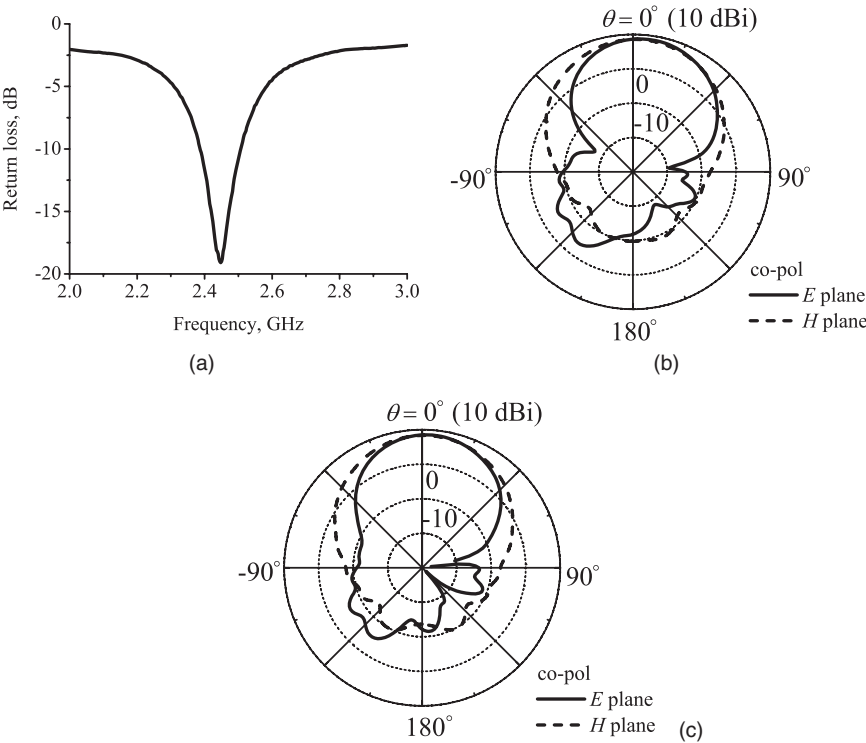


FIGURE 37-11 Measured results for the SPA applied in the WLAN AP applications (a) return loss; copolarized radiation patterns in *E* and *H* planes (b) at 2.4 GHz and (c) at 2.5 GHz

37.3 PLANAR MONOPOLE ANTENNAS

In this part, the state-of-the-art planar monopole antennas will be introduced as embeddable terminal antennas for mobile ultrawideband (UWB) devices, for instance, wireless universal serial bus (USB) dongles. First, the state-of-the-art planar monopoles are briefly summarized. Unique design criteria for UWB antenna design are described. After that, the planar monopoles applied to wireless USB are introduced.

The State of the Art

In general, UWB antennas indicate frequency-independent designs, for instance, TEM horn antennas, as shown in Figure 37-12a.²⁰⁻²² The antenna radiates linearly polarized TEM waves. Figure 37-12b shows a photo of a planar two-arm spiral antenna. The operating frequency range is determined by the inner and outer radii of the spiral. The radiated wave is a circularly polarized wave at the boresight. The spiral antenna is usually backed by a lossy or conducting cavity to improve impedance matching at low frequency and axial ratio by reducing the reflections from the end of the spiral arm. The phase center will change with operating frequency when a conical spiral antenna is used.²³ Bi-conical antennas may be the earliest antennas used in

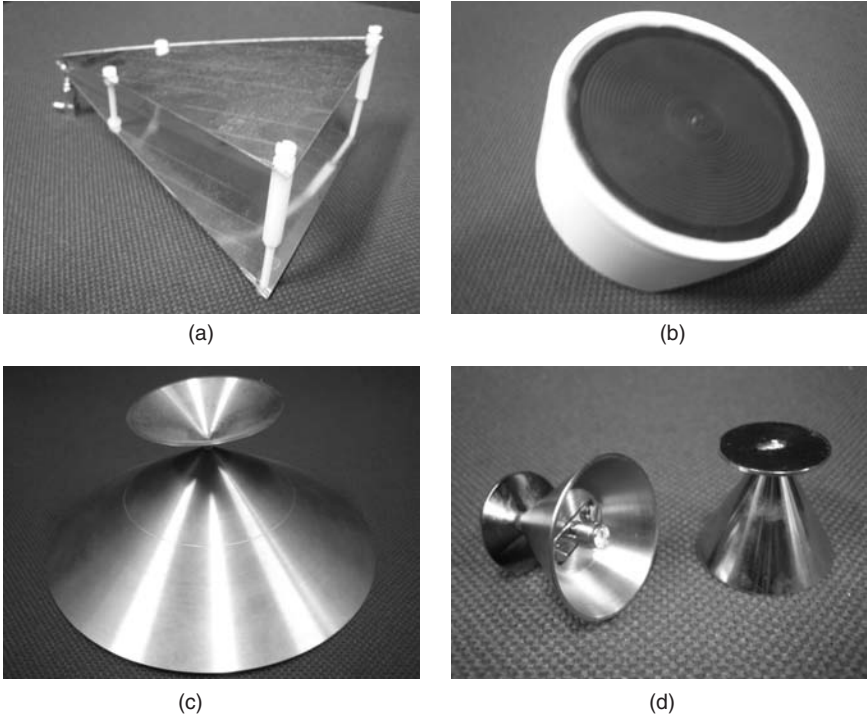


FIGURE 37-12 Frequency-independent antennas: (a) a TEM horn antenna; (b) a planar two-arm spiral antenna backed by a conducting cavity; (c) a bi-conical antenna; (d) two discone antennas

wireless communications because of their relatively stable phase centers and broad impedance bandwidths.²⁴ Figure 37-12c displays an asymmetrical finite bi-conical antenna with a broad impedance bandwidths.^{25–27} Figure 37-12d shows a typical discone antenna.

The existing frequency-independent antennas having a constant performance at all frequencies also include self-complementary log-periodic structures, such as planar log-periodic slot antennas, bidirectional log-periodic antennas, log-periodic dipole arrays, two- and four-arm log spiral antennas, and conical log-spiral antennas.²⁸

Furthermore, the cylindrical antennas can achieve broadband impedance characteristics by using resistive loading to form traveling wave along the dipole arms.^{29–31}

However, the antennas mentioned above are seldom used in portable/mobile devices due to the constraints of bulky size and high manufacturing cost, although they have been widely used in radar systems, electromagnetic compatibility (EMC), and channel measurements, as well as base stations for wireless communications. Therefore, planar monopoles or disc antennas have been alternatively used in wireless communications because they have shown excellent performance in impedance and radiation and because they have the significant advantage of small size.^{32–35} The earliest planar dipole may be the bow-tie antenna invented by Brown-Woodward.³⁶

Figure 37-13 reviews the planar trapezoidal designs.^{37–40} A general radiator with dimensions of W_u , W_b , H , and h that is installed above a large ground plane forms a planar monopole. The monopole antenna can be excited by a coaxial cable through a surface-mounted adapter (SMA) connector. The triangular monopole antenna is the case with $W_b = 0$ or $W_u = 0$, which can be considered as the planar versions of bi-conical antennas.

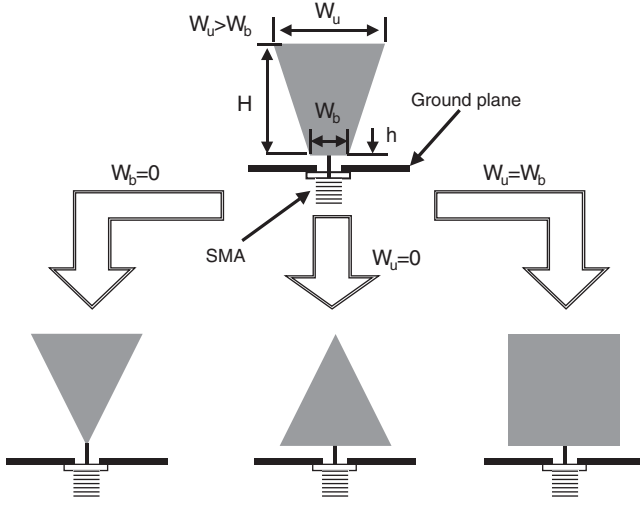


FIGURE 37-13 Review of the planar trapezoidal monopoles

The rectangular monopole antenna can be formed as $W_b = W_u \neq 0$, which is the planar version of cylindrical monopole antenna. The frequency corresponding to the lower edge of impedance bandwidth of the planar trapezoidal monopole can be estimated by Equations 37-2 and 37-3.

$$F(\text{GHz}) = \frac{0.25 \times 300}{H + h + \frac{W_u + W_b}{4\pi}} \quad (37-2)$$

$$F(\text{GHz}) = \frac{0.25 \times 300}{\sqrt{H^2 + \left(\frac{2 \max(W_u, W_b) - W_u - W_b}{2} \right)^2} + h + \frac{W_u}{2\pi}} \quad (37-3)$$

where

$$\max(W_u, W_b) = \begin{cases} W_u & W_u \geq W_b \\ W_b & W_b \geq W_u \end{cases}$$

and all dimensions are in millimeters. Equation 37-3 is the modified version of Eq. 37-2 to estimate the frequency corresponding to the lower edge of the impedance bandwidth and has been used to estimate the frequency corresponding to the lower edge of the impedance bandwidth of a rectangular radiator as a special case of polygonal radiators.

The rectangular planar monopole antenna can be derived from the trapezoidal planar antenna, as shown in Figure 37-13.⁴¹⁻⁴⁵ The monopole performance is mainly determined by the shape and size of the planar radiator as well as the feeding section. The overall size of the monopole and shape of the radiator dominate the frequency corresponding to the lower edge of the impedance bandwidth. The feed gap, the location of the feedpoint, and the shape of the radiator greatly affect the impedance matching.⁴⁶⁻⁴⁹ Figure 37-14 shows a rectangular planar monopole antenna and its modified versions. By modifying the shapes of the bottom sides of the radiators, the impedance matching of the planar rectangular monopole can be further improved, as shown in Figure 37-14b, c, and d, where the impedance transition

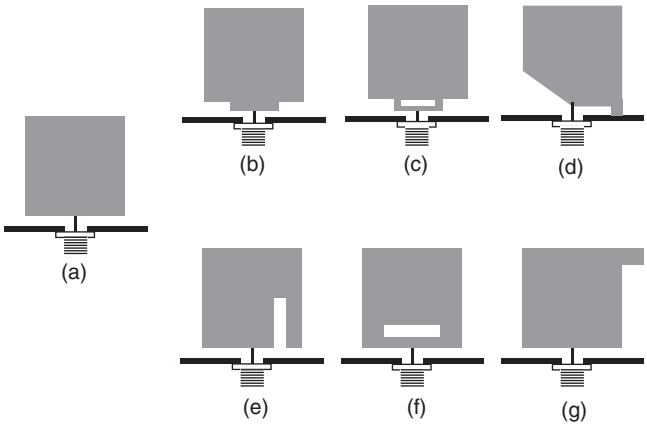


FIGURE 37-14 A rectangular planar monopole antenna and its modified versions

between the coaxial probe and the radiator has been improved by using step transition, two branch feeding strips, and a beveled radiator.^{46,47,49–51} A broadband impedance transition will ensure an impedance matching across a broad bandwidth. A shorting pin can also be used to make the planar antenna more compact.⁵² The impedance and radiation performance of four types of square planar monopole antennas were studied.⁴⁸

By slotting the radiators of the planar monopoles, the impedance bandwidth of the monopole can be improved where the current distributions on the radiators are changed, as shown in Figure 37-14e and f.^{53–55} By adding a horizontal strip, the height of the planar monopole can be reduced because of the longer effective current path, as shown in Figure 37-14g.⁵⁶

Theoretically, the radiators of the planar monopole antennas can be of any shape for broad operating bandwidth. The elliptical planar monopole antenna is one of the important types of planar monopoles in planar antenna design due to its broadband-even high-pass impedance performance. Figure 37-15 illustrates the elliptical monopoles in their general forms (Figure 37-15a), circular monopole (Figure 37-15b), semicircular monopole (Figure 37-15c), rectangular monopole with a semicircular bottom (Figure 37-15d), and annular monopole (Figure 37-15e).^{57–62}

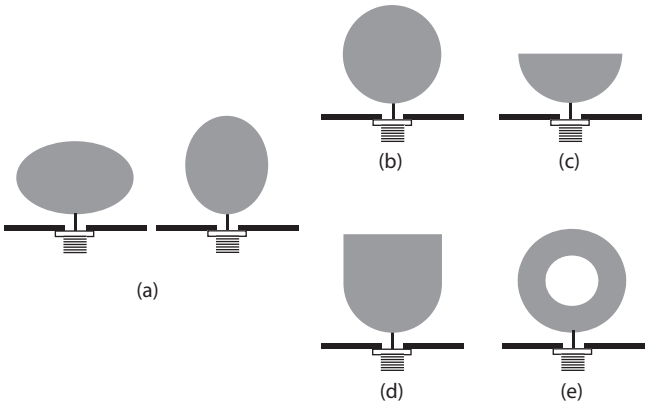


FIGURE 37-15 Elliptical planar antennas and their variations

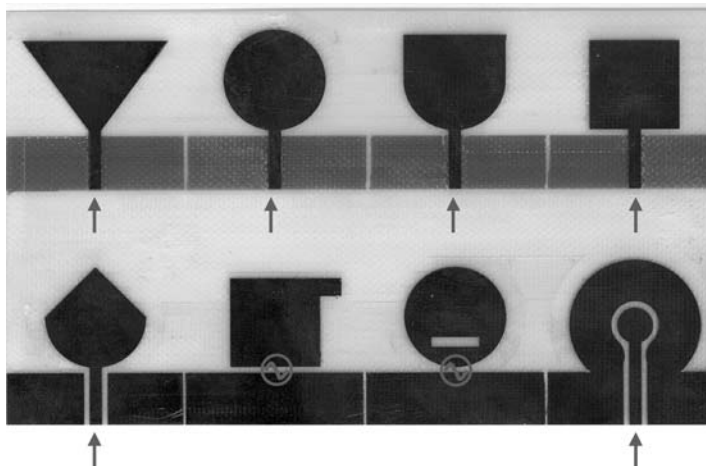


FIGURE 37-16 The prototypes of printed PCB monopoles: (a) a microstrip-fed triangular monopole; (b) a microstrip-fed circular PCB monopole; (c) a microstrip-fed rectangle+semicircle PCB monopole; (d) a microstrip-fed rectangular PCB monopole; (e) a CPW-fed triangle+semicircle PCB monopole; (f) a directly cable-fed rectangular PCB monopole with a strip; (g) a directly cable-fed circular PCB monopole with a slot; (h) a CPW-fed circular slot PCB antenna (from left to right and from the first to the second row).

However, in mobile wireless applications, the antennas are expected to be embeddable or easy to be integrated into wireless devices in system design. Therefore, the antennas directly printed onto a printed circuit board (PCB) are the most promising designs. Such antennas are usually constructed by etching the radiators onto the dielectric substrate of PCB stabs and a ground plane around the radiators.

Figure 37-16 demonstrates some prototypes of PCB antennas. It can be seen that the PCB antennas can be fed by a microstrip transmission line (Figure 37-16a–d), a coplanar waveguide (CPW) structure (Figure 37-16e and h), or directly by an RF cable where the inner conductor is connected to the upper radiator and the outer conductor is grounded into a system ground (Figure 37-16f and g), as shown in Figure 37-16.

For the microstrip-fed cases, the ground planes are usually etched on opposite surfaces, which are connected into the system ground. The ground planes are etched close to the feeding strip on the same surface as the feeding strip for the CPW-fed PCB monopoles. The directly cable-fed monopoles are often used in scenarios where the PCB monopoles are not close to the main board and an RF cable is used to connect the antenna into the main board—for example, the antennas embedded into the cover of a laptop computer.

Antennas Applied in High-Speed UWB Wireless Communications

UWB is a technology for the transmission of information by using an extremely large operating bandwidth. Since the first Report and Order by the Federal Communications Commission (FCC) released the unlicensed use of UWB on February 14, 2002, the UWB technology has been expected to be the next-generation technology for short-range but high-speed wireless communications.⁶³ The frequency range of 3.1–10.6 GHz has been allocated for such applications. The emission limit masks are regulated by regulators such

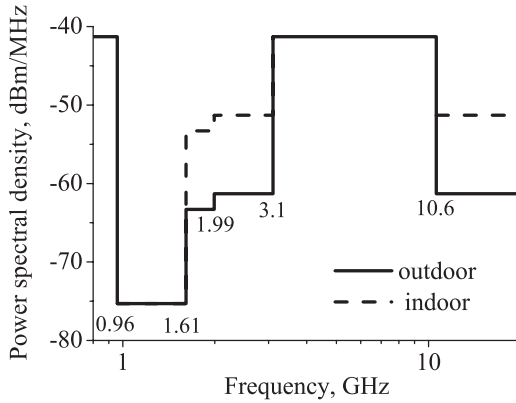


FIGURE 37-17 Emission limit masks for indoor and outdoor UWB applications

as the FCC for co-existence with other electronic systems, as shown in Figure 37-17. The emission power limits are lower than the noise floor in order to avoid possible interference between UWB devices and existing electronic systems. The masks may vary against regions or countries but the maximum emission levels are always kept lower than -41.3 dBm/MHz.

The extremely large spectrum provides room to use extremely short pulses on the order of picoseconds. Thus, the pulse repetition or data rates can be low or very high, typically up to giga pulse/second. The pulse rates are dependent on the applications.

Due to the unique requirements of UWB systems, the antenna design faces many challenges. Besides the requirements for conventional broadband antennas in terms of impedance bandwidth, gain, radiation patterns, and polarization, the special requirements for the UWB antenna in general include the following two points.

First, both the frequency and time domain responses of the UWB antenna should be taken into account due to the extremely wide bandwidth and direct pulsed signals. The frequency domain response includes all conventional parameters such as impedance, radiation, and transmission. The impedance bandwidth is measured in terms of return loss or voltage standing wave ratio (VSWR). Usually, the return loss should be less than -10 dB or $VSWR < 2:1$. An antenna with an impedance bandwidth narrower than the operating bandwidth tailors the spectrum of transmitted and received signals, acting as a bandpass filter in the frequency domain, and reshapes the waveforms of radiated or received pulses in the time domain. The radiation performance includes radiation efficiency, radiation patterns, polarization, and gain. The radiation efficiency is an important parameter especially for small antenna design, where it is difficult to achieve impedance matching due to small radiation resistance and large reactance. For the small antenna with weak radiation directivity, the radiation efficiency is of greater practical interest than the gain. The radiation patterns show the directions where the signals will be transmitted.

Second, the effect of the modulation schemes of the systems on the antenna design should be taken into consideration. The wireless communication systems based on UWB technology can be OFDM based or pulsed based. In an OFDM-based system, the whole UWB band is occupied by many sub-bands, where each sub-band having a few hundred megahertz (larger than 500 MHz) can be considered as broadband. Within the sub-bands, the effect of nonlinearity of the phase shift on the receiving performance can be ignored because the phase varies very slowly with frequency. Therefore, the design of the antenna is more focused on achieving constant frequency response in terms of the radiation efficiency,

gain, return loss, radiation patterns, and polarization over the operating band, which may fully or partially cover the UWB bandwidth of 7.5 GHz.

The pulse-based UWB systems usually require linear phase, impedance, and gain responses that can entirely cover the operating bandwidth or partially cover the bandwidth where the majority of the pulse energy is distributed. To prevent the distortion of the received pulses, a UWB antenna that can produce radiation fields of constant magnitude and a phase shift that varies linearly with frequency is desired.

In brief, an OFDM-based UWB system has almost the same requirements as those in a broadband system but with an extremely broad bandwidth, which usually varies from 50 percent (for the lower UWB band of 3.1–5 GHz) to 100 percent (for the entire UWB band of 3.1–10.6 GHz). However, paying additional attention to pulse-based UWB systems is needed, where a UWB antenna usually acts as a bandpass filter that reshapes the spectra of the radiated/received pulses. To design a good UWB antenna, we should keep the following important design factors in mind:

- Enough impedance bandwidth to cover the operating bandwidth
- Steady directional or omnidirectional gain radiation patterns
- High and constant gain along desired direction(s)
- Linear phase, especially for pulse-based systems
- Consistent polarization
- Mobility with small size/low profile
- Low design complicity
- Low material/manufacturing cost

Based on these requirements, a circular planar dipole antenna is exemplified. Figure 37-18 shows a pair of the circular planar dipoles, which are face-to-face positioned for transfer function (S_{21}) measurement. The planar circle has a radius of 10.5 mm and feed gap of 2.2 mm. The identical antennas are used as transmit and receive antennas in the test setup shown in Figure 37-18, a pair of the planar dipole antennas with a separation of 100 mm and positioned in parallel and face-to-face.

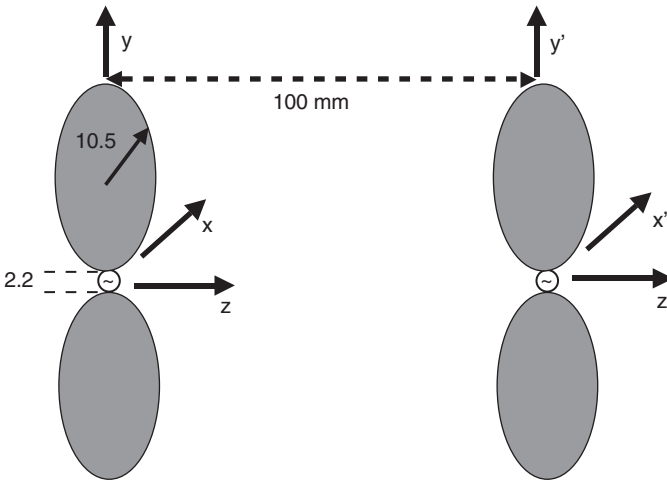


FIGURE 37-18 A pair of circular planar dipole antennas

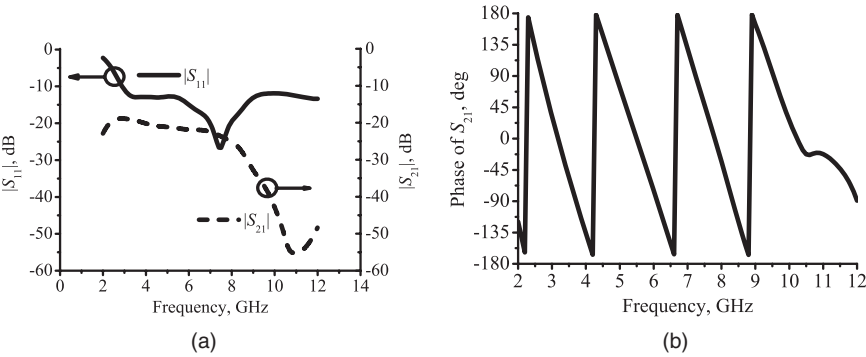


FIGURE 37-19 (a) Return loss $|S_{11}|$ and magnitude of $|S_{21}|$; (b) the phase response of $|S_{21}|$ of the circular dipole antenna

Figure 37-19 demonstrates the return losses $|S_{11}|$ and the magnitude of the transfer function $|S_{21}|$ for the dipole shown in Figure 37-18. The circular planar dipole antenna demonstrates the broad impedance bandwidth covering the entire UWB band and broad transmission coverage 2–8.5 GHz within a 10-dB variation. The results shown in Figure 37-19 are simulated.

Antennas in UWB-based Wireless USB

The most promising applications of UWB technology in the short-range and high-speed wireless interfaces may be wireless USB (WUSB), which replaces wired USB to match the data rate of USB 2.0 at 480 Mbps. The technology is a hub-and-spoke connection that supports dual-role devices in which a product such as a camera can act either as a device to a host laptop/desktop or as a host to a device such as a printer. In this section, a printed PCB antenna will be designed for a WUSB dongle, which is used in a laptop environment.

Figure 37-20a shows a printed PCB antenna on a dielectric substrate with a popular USB flash drive. The radiator and ground plane are etched on opposite sides of the PCB, as shown in Figure 37-20b. The radiator consists of a rectangular section and a horizontal

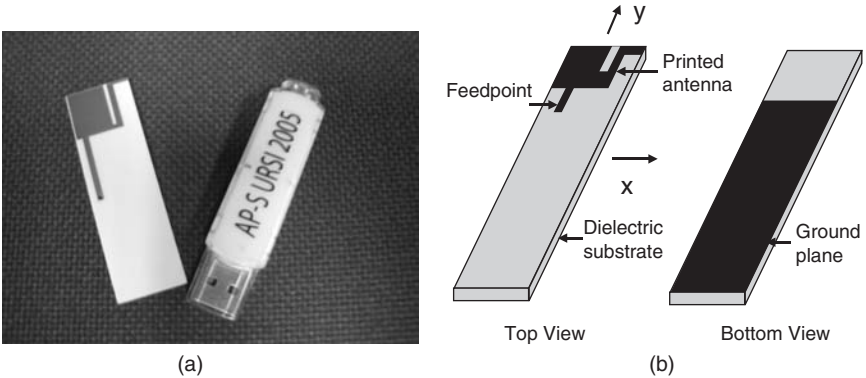


FIGURE 37-20 A small PCB antenna for a WUSB dongle

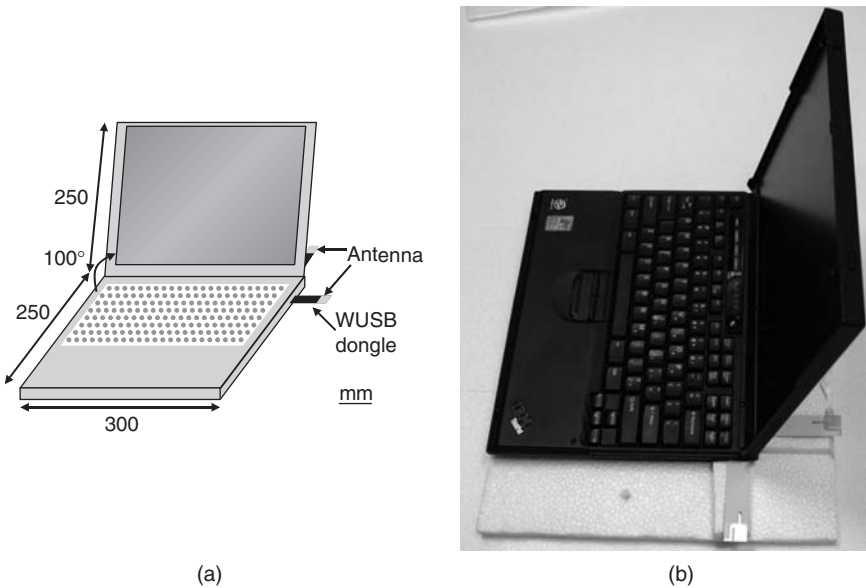


FIGURE 32-21 (a) Placement of the antennas; (b) the antennas under test

strip, which is used to reduce the length of the rectangular radiator.⁵⁶ A rectangular notch was cut close to the horizontal strip.⁶⁴ The effects of the ground plane on the performance of the printed PCB antenna have greatly been reduced because of the introduction of the notch. The radiator is fed by a microstrip line from the left side of the radiator. The excitation is located at the bottom end of the microstrip line. The gap between the radiator and the ground plane is typically around 1 mm. The ground plane has a size of a typical wireless USB dongle, where the system will be installed.

A WUSB dongle is placed beside a laptop. The laptop usually has a metallic and lossy cover and main body. To reduce the effects of the cover and body, the antenna is installed as far as possible from the cover and main body. As an example, the effect of the laptop on the radiation performance of the antenna is investigated. Figure 37-21 shows a dongle with a small printed PCB antenna installed at an IBM ThinkPad laptop under test. The keyboard panel of the laptop measures $250 \times 300 \times 25$ mm. The screen measures 250×300 mm and is opened as if it is in use such that the angle between the screen and the keyboard panel is 100° . The antenna can be placed at different positions. Here, two different positions on the laptop for analysis are shown in Figure 37-21. The antenna can be placed at the right side of the back panel P1 and at the edge on the right-side panel P2. The size of the ground plane of the antenna does not affect the impedance and radiation performance significantly because the effect of the ground plane on the antenna performance has been significantly reduced.⁶⁴ Therefore, no matter if the dongle is electrically connected to the metal casing or not, the performance of the antenna is unchanged.

Figure 37-22 plots the radiation patterns for the total fields in the horizontal (x - y) plane when the PCB antenna is placed at two positions on the laptop and in free space. The coordinate system is shown in Figure 37-20. By comparing the gain for the antenna in free space and at positions P1 and P2, it can be seen that the antenna at P1 has higher gain than that at P2 due to the blockage effect of the cover on the radiation from the antenna in

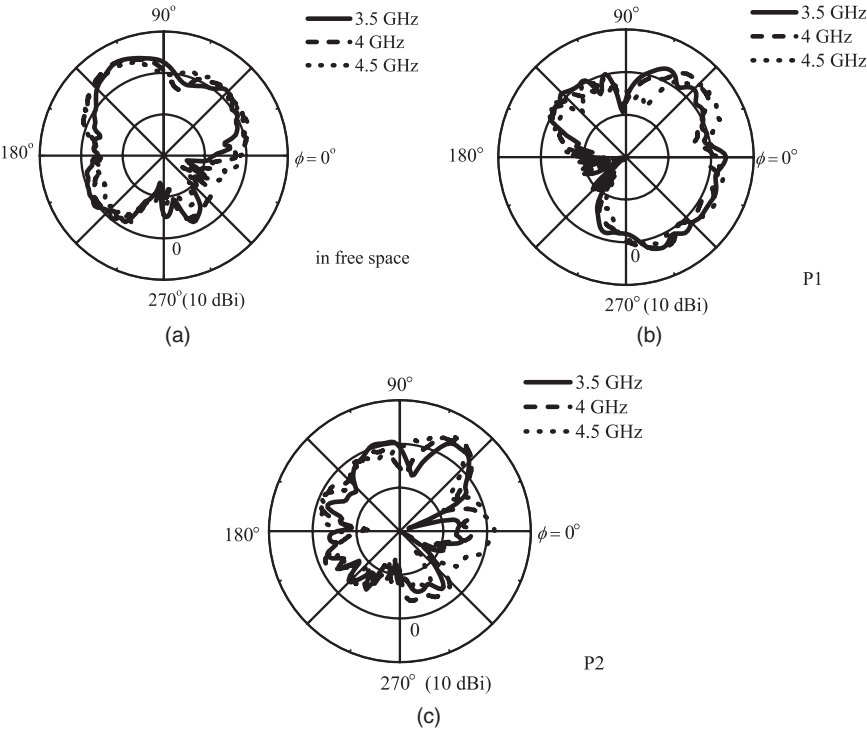


FIGURE 37-22 Measured gain patterns for total fields in the x - y plane (a) in free space and (b) for the antenna positions P1 and (c) P2 on the laptop

proximity of the laptop. The reflection from the cover leads to the stronger radiation from the antenna at P1 in the span of 160° – 90° – 0° – 270° than that within the rest range. The radiation for the antenna at P2 is nearly omnidirectional but with several nulls and lower gain than that for the antenna at P1 in most of the directions, although the blockage of the cover on the antenna at P2 should be less than that on the antenna at P1.

In short, UWB is a promising technology for short-range high-speed wireless communications. To accommodate the special requirements for the antennas, the traits of UWB systems should be studied carefully and effects of the installation environments on the antenna performance should be taken into account in the system's point of view, in particular for small and embedded antennas.

REFERENCES

1. Z. N. Chen and M. Y. W. Chia, *Broadband Planar Antennas: Design and Applications* (New York: John Wiley & Sons, Inc. 2006).
2. J. Howell, "Microstrip Antennas," *Antennas and Propagation Society International Symposium*, vol. 10 (December 1972): 177–180.
3. R. Munson, "Conformal Microstrip Antennas and Microstrip Phased Arrays," *IEEE Trans. on Antennas and Propagat.*, vol. 22, no. 1 (January 1974): 74–78.

4. D. M. Pozar, "A Microstrip Antenna Aperture Coupled to a Microstrip Line," *Electronics Letters*, vol. 21 (January 17, 1985): 49–50.
5. J. M. Griffin and J. R. Forrest, "Broadband Circular Disc Microstrip Antenna," *Electronics Letters*, vol. 18, no. 5 (March 1982): 266–269.
6. K. F. Lee, K. Ho, and J. Dahele, "Circular-Disk Microstrip Antenna with an Air Gap," *IEEE Trans. on Antennas and Propagat.*, vol. 32, no. 8 (August 1984): 880–884.
7. K. S. Fong, H. F. Guse, and M. J. Withers, "Wideband Multiplayer Coaxial-Fed Microstrip Antenna Element," *Electronics Letters*, vol. 21, no. 11 (May 1985): 497–499.
8. G. Mayhew-Ridgers, J. W. Odendaal, and J. Joubert, "Single-Layer Capacitive Feed for Wideband Probe-Fed Microstrip Antenna Elements," *IEEE Trans. on Antennas and Propagat.*, vol. 51, no. 6 (June 2003): 1405–1407.
9. T. Huynh and K. F. Lee, "Single-Layer Single-Patch Wideband Microstrip Antenna," *Electronic Letters*, vol. 31, no. 16 (August 1995): 1310–1312.
10. Z. N. Chen, "Experimental Investigation on Rectangular Plate Antenna with Ω -shaped Slot," *Radio Science*, vol. 36, no. 5 (August–September 2001): 833–840.
11. H. Nakano, M. Yamazaki, and J. Yamauchi, "Electromagnetically Coupled Curl Antenna," *Electronic Letters*, vol. 33, no. 12 (June 1997): 1003–1004.
12. K. M. Luk, C. L. Mak, Y. L. Chow, and K. F. Lee, "Broadband Microstrip Antenna," *Electronic Letters*, vol. 34, no. 15 (July 1998): 1442–1443.
13. Z. N. Chen and M. Y. W. Chia, "A Feeding Scheme for Enhancing Impedance Bandwidth of a Suspended Plate Antenna," *Microwave and Optical Technology Letters*, vol. 38, no. 1 (2003): 21–25.
14. P. S. Hall, "Probe Compensation in Thick Microstrip Patches," *Electronic Letters*, vol. 23 (1987): 606–607.
15. L. Levis, A. Ittipiboon, and A. Petosa, "Probe Radiation Cancellation in Wideband Probe-Fed Microstrip Arrays," *Electronic Letters*, vol. 36 (2000): 606–607.
16. Z. N. Chen and M. Y. W. Chia, "Broadband Suspended Probe-Fed Plate Antenna with Low Cross-Polarization Levels," *IEEE Trans. on Antennas and Propagat.*, vol. 51, no. 2 (February 2003): 345–346.
17. Z. N. Chen and M. Y. W. Chia, "A Novel Center-Fed Suspended Plate Antenna," *IEEE Trans. on Antennas and Propagat.*, vol. 51, no. 6 (June 2003): 1407–1410.
18. Z. N. Chen and M. Y. W. Chia, "Center-Fed Microstrip Patch Antenna," *IEEE Trans. Antennas and Propagat.*, vol. 51, no. 3 (March 2003): 483–487.
19. Z. N. Chen, and M. Y. W. Chia, "Broadband Suspended Plate Antennas Fed by Double L-shaped Strips," *IEEE Trans. on Antennas and Propagat.*, vol. 52, no. 9 (September 2004): 2496–2500.
20. M. Kanda, "Transients in a Resistively Loaded Linear Antennas Compared with Those in a Conical Antenna and a TEM Horn," *IEEE Trans. on Antennas and Propagat.*, vol. 28, no. 1 (January 1980): 132–136.
21. L. T. Chang and W. D. Burnside, "An Ultrawide-Bandwidth Tapered Resistive TEM Horn Antenna," *IEEE Trans. on Antennas and Propagat.*, vol. 48, no. 12 (December 2000): 1848–1857.
22. R. T. Lee and G. S. Smith, "On the Characteristic Impedance of the TEM Horn Antenna," *IEEE Trans. on Antennas and Propagat.*, vol. 52, no. 1 (January 2004): 315–318.
23. T. W. Hertel and G. S. Smith, "On the Dispersive Properties of the Conical Spiral Antenna and Its Use for Pulsed Radiation," *IEEE Trans. on Antennas and Propagat.*, vol. 51, no. 7 (July 2003): 1426–1433.
24. J. D. Kraus, *Antennas*, 2nd Ed. (New York: McGraw-Hill, 1988): 340–358.
25. C. W. Harrison, Jr. and C. S. Williams, Jr., "Transients in Wide-Angle Conical Antennas," *IEEE Trans. on Antennas and Propagat.*, vol. 13, no. 2 (March 1965): 236–246.
26. S. S. Sandler and R. W. P. King, "Compact Conical Antennas for Wide-band Coverage," *IEEE Trans. on Antennas and Propagat.*, vol. 42, no. 3 (March 1994): 436–439.
27. S. N. Samaddar and E. L. Mokole, "Biconical Antennas with Unequal Cone Angles," *IEEE Trans. on Antennas and Propagat.*, vol. 46, no. 2 (February 1998): 181–193.
28. P. E. Mayes, "Frequency-Independent Antennas and Broad-band Derivatives Thereof," *Proceedings of the IEEE*, vol. 80, no. 1 (January 1992): 103–112.

29. T. T. Wu and R. W. P. King, "The Cylindrical Antenna with Nonreflecting Resistive Loading," *IEEE Trans. on Antennas and Propagat.*, vol. 13, no. 3 (May 1965): 369–373.
30. D. L. Senguta and Y. P. Liu, "Analytical Investigation of Waveforms Radiated by a Resistively Loaded Linear Antenna Excited by a Gaussian Pulse," *Radio Science*, vol. 9, no. 3 (June 1974): 621–630.
31. J. G. Maloney and G. S. Smith, "A Study of Transient Radiation from the Wu-King Resistive Monopole—FDTD Analysis and Experimental Measurements," *IEEE Trans. on Antennas and Propagat.*, vol. 41, no. 5 (May 1993): 668–676.
32. H. Meinke and F. W. Gundlach, *Taschenbuch der Hochfrequenztechnik* (Berlin: Springer-Verlag, 1968): 531–535.
33. G. Dubost and Zisler, *Antennas a Large Bande* (Paris: Masson, 1976): 128–129.
34. S. Honda, M. Ito, H. Seki, and Y. Jinbo, "A Disk Monopole Antenna with 1:8 Impedance Bandwidth and Omnidirectional Radiation Pattern," *Int. Symp. Antennas Propagat.* (1992): 1145–1148.
35. M. Hammoud, P. Poey, and F. Colombel, "Matching the Input Impedance of a Broadband Disc Monopole," *Electronics Lett.*, vol. 29 (1993): 406–407.
36. G. H. Brown and O. M. Woodward, "Experimentally Determined Radiation Characteristics of Conical and Triangular Antennas," *RCA Review*, vol. 13 (December 1952): 425–452.
37. Z. N. Chen and M. Y. W. Chia, "Impedance Characteristics of Trapezoidal Planar Monopole Antenna," *Microw. Opt. Techno. Lett.*, vol. 27, no. 2 (October 2000): 120–122.
38. J. A. Evans and M. J. Ammann, "Planar Trapezoidal and Pentagonal Monopoles with Impedance Bandwidths in Excess of 10:1," *IEEE Antennas and Propagation Society International Symposium*, vol. 3 (July 11–16, 1999): 1558–1561.
39. X. H. Wu, Z. N. Chen, and N. Yang, "Optimization of Planar Diamond Antenna for Single/Multiband UWB Wireless Communications," *Microw. Opt. Techno. Lett.*, vol. 42, no. 6 (2004): 451–455.
40. Z. N. Chen, "Impedance Characteristics of Planar Bow-tie-like Monopole Antennas," *Electronics Letters*, vol. 36, no. 13 (2000): 1100–1101.
41. M. J. Ammann, "Square Planar Monopole Antenna," *IEE National Conference on Antennas and Propagation* (March 31–April 1, 1999): 37–40.
42. M. J. Ammann, "Impedance Bandwidth of the Square Planar Monopole," *Microw. Opt. Techno. Lett.*, vol. 24 (2000): 185–187.
43. M. J. Ammann and Z. N. Chen, "An Asymmetrical Feed Arrangement for Improved Impedance Bandwidth of Planar Monopole Antennas," *Microw. Opt. Techn. Lett.*, vol. 40, no. 2 (2004): 156–158.
44. Z. N. Chen, M. J. Ammann, and M. Y. W. Chia, "Broadband Square Annular Planar Monopoles," *Microw. Opt. Techno. Lett.*, vol. 36, no. 6 (March 2003): 449–454.
45. Z. N. Chen, "Experimental on Input Impedance of Tilted Planar Monopole Antennas," *Microw. Opt. Techno. Lett.*, vol. 26, no. 3 (2000): 202–204.
46. K. G. Thomas, N. Lenin, and R. Sivaramakrishnan, "Ultrawideband Planar Disc Monopole," *IEEE Trans. on Antennas and Propagat.*, vol. 54, no. 4 (April 2006): 1339–1341.
47. S. Su, K. Wong, and C. Tang, "Ultrawideband Square Planar Antenna for IEEE 802.16a Operating in the 2–11 GHz Band," *Microw. Opt. Techn. Lett.*, vol. 42, no. 6 (September 2004): 463–466.
48. X. H. Wu and Z. N. Chen, "Comparison of Planar Dipoles in UWB Applications," *IEEE Trans. on Antennas and Propagat.*, vol. 53, no. 6 (2005): 1973–1983.
49. M. J. Ammann and Z. N. Chen, "A Wideband Shorted Planar Monopole with Bevel," *IEEE Trans. on Antennas and Propagat.*, vol. 51, no. 4 (2003): 901–903.
50. X. H. Wu, A. A. Kishk, and Z. N. Chen, "A Linear Antenna Array for UWB Applications" *IEEE Int. Symp. Antennas and Propagat.*, vol. 1A (July 3–8, 2005): 594–597.
51. E. Antonino-Daviu, M. Cabedo-Fabres, M. Ferrando-Bataller, and A. Valero-Nogueira, "Wideband Double-Fed Planar Monopole Antennas," *Electronics Letters*, vol. 39, no. 23 (November 2003): 1635–1636.

52. E. Lee, P. S. Hall, and P. Gardner, "Compact Wideband Planar Monopole Antenna," *Electronics Letters*, vol. 35, no. 25 (December 1999): 2157–2158.
53. H. S. Choi, J. K. Park, S. K. Kim, and J. Y. Park, "A New Ultra-Wideband Antenna for UWB Applications," *Microw. Opt. Techno. Lett.*, vol. 40, no. 11 (May 2004): 399–401.
54. S. Y. Suh, W. L. Stutzman, and W. A. Davis, "A New Ultrawideband Printed Monopole Antenna: the Planar Inverted Cone Antenna (PICA)," *IEEE Trans. on Antennas and Propagat.*, vol. 52, no. 5 (May 2004): 1361–1364.
55. D. Valderas, J. Meléndez, and I. Sancho, "Some Design Criteria for UWB Planar Monopole Antennas: Application to a Slotted Rectangular Monopole," *Microw. Opt. Techno. Lett.*, vol. 46, no. 1 (July 2005): 6–11.
56. A. Cai, T. S. P. See, and Z. N. Chen, "Study of Human Head Effects on UWB Antenna," *IEEE Intl. Workshop on Antenna Technology* (March 7–9, 2005): 310–313.
57. Z. N. Chen, M. J. Ammann, M. Y. W. Chia, and T. S. P. See, "Circular Annular Planar Monopoles with EM Coupling," *IEE Proc. Microw. Antennas, Propagat.*, vol. 150, no. 4 (August 2003): 269–273.
58. N. P. Agrawall, G. Kumar, and K. P. Ray, "Wide-band Planar Monopole Antenna," *IEEE Trans. on Antennas and Propagat.*, vol. 46, no. 2 (February 1998): 294–295.
59. T. Yang and W. A. Davis, "Planar Half-Disk Antenna Structures for Ultra-Wideband Communications," *IEEE Int. Symp. Antennas and Propagat.*, vol. 3 (June 2004): 2508–2511.
60. C. Y. Huang and W. C. Hsia, "Planar Elliptical Antenna for Ultra-Wideband Communications," *Electronics Letters*, vol. 41, no. 6 (March 2005): 296–297.
61. P. V. Anob, K. P. Ray, and G. Kumar, "Wideband Orthogonal Square Monopole Antennas with Semi-Circular Base," *IEEE Int. Symp. Antennas and Propagat.*, vol. 3 (July 2001): 294–297.
62. J. W. Lee, C. S. Cho, and J. Kim, "A New Vertical Half Disc-Loaded Ultra-Wideband Monopole Antenna (VHDM) with a Horizontally Top-Loaded Small Disc," *Antennas and Wireless Propagat. Letters*, vol. 4 (2005): 198–201.
63. *First Report and Order, Federal Communications Commission (FCC)*, February 14, 2002.
64. Z. N. Chen, T. S. P. See, and X. M. Qing, "Small Ground-Independent Planar UWB Antenna," *IEEE Int. Symp. Antennas and Propagat.*, Albuquerque, New Mexico, July 2006.

Chapter 38

Antennas for Medical Applications

Cynthia M. Furse

University of Utah

CONTENTS

38.1	OVERVIEW OF ANTENNAS FOR MEDICAL APPLICATIONS	38-2
38.2	THE ENVIRONMENT	38-2
38.3	ANTENNAS FOR MEDICAL IMAGING	38-6
38.4	HEATING	38-11
38.5	COMMUNICATION (BIOTELEMETRY)	38-13
38.6	PULSED ELECTROMAGNETIC FIELDS	38-16
38.7	SENSING	38-17
38.8	FUTURE DIRECTIONS	38-19

38.1 OVERVIEW OF ANTENNAS FOR MEDICAL APPLICATIONS

Antennas used for medical applications span applications in imaging, communication with implantable medical devices, heating for treatment of cancer, cardiac abnormalities, and hypothermia, measurement of fields for assessment of RF safety, augmentation of healing, and reduction of pain. Some of these applications have gained worldwide acceptance and are currently used with human subjects, and others are still in the research and development stage. This chapter describes the nature of the human body environment in which these antennas are commonly used, relevant regulations and guidelines, and the antennas and their applications.

38.2 THE ENVIRONMENT

The Regulatory Environment

There are two types of regulations of particular interest to designers of antennas for medical applications. The first is the allowable frequency. Applications that are used external to the body or for short periods of time (hyperthermia treatment, pain control, cardiac ablation, etc.) utilize the Industrial, Scientific, and Medical (ISM) bands (433, 915, 2450 MHz) in both the United States and Europe. Higher frequencies have the advantage of smaller antenna sizes, but the disadvantage of lower depths of penetration within the body. Implantable medical devices that are meant to stay in the body for a long period of time have been allocated a band of their own in the United States, the Medical Implant Communication Service (MICS) band from 402–405 MHz.¹ The maximum bandwidth that can be used by a single device is 300 kHz in this band. The maximum power limit is 25 μ W Equivalent Radiated Power (ERP).^{2,3} MICS shares its frequency allocation with the Meteorological Aids Service (METAIDS), which is used primarily by weather balloons, and is therefore specified for indoor use.

The second type of regulation of interest in the design of antennas for medical applications is the limit on allowable absorbed power in the body. The limits for whole-body exposure are generally not the limiting factor. Instead, limits on localized power are more critical. Localized power is defined by specific absorption rate (SAR), which is calculated as

$$\text{SAR (W/kg)} = \frac{\sigma |\bar{E}_p|^2}{2\delta} \quad (38-1)$$

where σ is the electrical conductivity of the tissue (S/m), \bar{E}_p is the peak value of the electric field, and δ is the density of the tissue (kg/m^3). Localized SAR for nontherapeutic applications is limited to 2 W/kg in any 10-gram region of the body with an approximately cubic volume.^{4,5,6} An important exception is made in the new IEEE Safety Standard⁵ for the pinna (the outer ear) for an increased limit of 4.0 W/kg for the general public and 20.0 W/kg for occupational exposures. For therapeutic applications such as cardiac ablation and hyperthermia, the absorption limit does not apply, and care must then be taken not to damage surrounding tissues by overheating them.

The question of whether electromagnetic radiation causes harm to the body at nonthermal levels has been hotly contested and remains a topic of ongoing research. A focused

review of this research was completed by the NIEHS in 1999.⁷ Significant research and ongoing professional oversight have lead to today's RF exposure standards, which are meant to ensure safe levels of RF fields.⁸ Still, designers of antennas for medical systems should be aware of public fears that may accompany their use and should anticipate being asked to address this question.

The Physical Environment

Antennas used for medical applications are strongly impacted by the lossy dielectric materials that make up the human body. This presents challenges when antennas are used for a communication system, because the tissues absorb the power and detune the antenna. This depends strongly on the frequency and the location and depth of the antenna in or near the body, and significant variation can be seen from patient to patient. Figure 38-1 shows the electrical properties of muscle and fat as a function of frequency.⁹ At low frequencies, the conductivity of the tissue dominates the impact on the field, and at high frequencies, the dielectric values tend to dominate. Table 38-1⁹ shows the electrical properties of several different tissues in the body at 433 MHz, which is a commonly used ISM frequency. Muscle is highly conductive and therefore very lossy, whereas fat has lower conductivity and therefore lower loss. These two tissues are near extremes in the body. A common rough approximation is that the body can be modeled using average properties of 2/3 muscle. This is suitable for addressing global questions such as total power absorbed in the body, but is generally not suitable for evaluating near field effects such as peak SAR.

The lossy tissues in the body have several effects on antennas used for medical applications. When antennas are used for deliberately depositing power in the body for hyperthermia or cardiac ablation, for instance, the power tends to stay more localized around the antenna, where it is absorbed and converted to heat. For cardiac ablation or in vitro hyperthermia applicators, this is good, because it means that the heat will not penetrate to nearby structures that are not meant to be heated. For external, whole-body applicators for hyperthermia or for medical imaging applications, this loss means that it can be difficult to get the power to penetrate deep within the body. Lower frequencies are used when possible, and regions near the surface of the body (such as the breast) are easier to work with than areas deep within the torso. Multiple antennas must be used outside the body and focused in some way in order to get the power deep into the body.

For biotelemetry (communication) applications, the same types of problems are seen. Communication with subcutaneous implants loses less power in the body than communication with deep body implants, for instance. Power lost in the body has two effects: it is wasted and cannot be used for communication, and the RF exposure limits typically limit

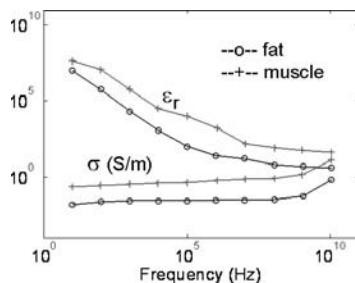


FIGURE 38-1 Electrical properties of muscle and fat (from Gabriel⁹)

TABLE 38-1 Electrical Properties of Tissues at 433 MHz (from Gabriel⁹)

Tissue	ϵ_r	σ (S/m)
Aorta	49.15	0.7395
Bladder	17.67	0.3128
Blood	57.3	1.72
Bone (Cancellous)	21.08	0.02275
Bone (Cortical)	13.77	0.1032
Bone (Marrow)	5.137	0.03575
Breast Fat	5.62	0.04953
Cartilage	43.64	0.65
Cerebellum	52.9	0.91
Cerebro Spinal Fluid	68.97	2.32
Cervix	44.17	1.020
Colon	60.88	0.96
Cornea	54.4	1.070
Dura	51.03	0.8
Eye Tissues	57.69	1.010
Fat	5.028	0.04502
Gall Bladder	60.06	1.035
Gall Bladder Bile	76.55	1.613
Grey Matter	54.27	0.8775
Heart	60.74	0.9866
Kidney	57.3	1.152
Lens Cortex	52.75	0.6742
Lens Nucleus	38.76	0.38
Liver	50.34	0.68
Lung Deflated	52.83	0.7147
Lung Inflated	21.58	0.3561
Muscle	64.21	0.9695
Nerve	35.7	0.500
Ovary	51.55	1.033
Skin (dry)	42.48	0.5495
Skin (wet)	51.31	0.72
Small Intestine	74.1	2.053
Spleen	60.62	1.041
Stomach	74.55	1.120
Tendon	50.53	0.7554

TABLE 38-1 *Continued*

Tissue	ϵ_r	σ (S/m)
Testes	65.2	1.137
Thyroid	60.02	0.8183
Tongue	58.79	0.8993
Tracea	42.93	0.673
Uterus	64.73	1.117
Vitreous Humour	66.16	0.3931
White Matter	39.84	0.5339

the power that can be used for communication and hence the range or bandwidth of the system. Also, the lower the frequency used (to enable penetration and minimize power loss), the larger the antenna must be. Determining the exact impact of the body on the antenna generally requires precise calculation of the antenna in the presence of the human body. This typically requires numerical methods that are described in the next section.

Numerical Simulation of Antennas in or near the Human Body

Several methods for analyzing antenna arrays for medical applications exist.¹⁰ For simple cases where the biological structure can be approximated as uniform or by very simple models such as layers or cylinders, classical methods such as analytical analysis^{11,12} or method of moments^{13,14} can be used. If the structure of the body varies so much that anatomically precise modeling is rendered imprecise by variation between individuals, these simple analyses can be used to determine an optimal array design for the range of expected variation between individuals. An example of this was done in Hadley¹⁵ for design of coils for vascular MRI. Another example in which the body can be modeled as near-uniform is in the case of arrays for hyperthermia of the brain. In Furse and Iskander¹⁶ stepped-impedance dipoles were modeled using method of moments in an homogenous brain with a localized (nonhomogenous) tumor. Method of moments with a simple pulsed basis function (which is the most numerically efficient form) has limitations for heterogeneous models, however, due to artificial charge buildup on the dielectric interfaces.¹³ Higher-order basis functions can overcome this limitation, although the computational complexity is significantly increased.¹⁷ In addition, the method of moments is very computationally expensive when heterogeneous models are evaluated. It requires $N \log N$ computations, where N is the number of cells in the model, including those making up the heterogeneous object.

A more efficient method for calculation of heterogeneous objects is the finite difference time domain (FDTD) method, which has led to its tremendous popularity for numerical bioelectromagnetic calculations. For example, a interstitial array of hyperthermia applicators simulated using method of moments¹⁶ was simulated with a fraction of the computational resources using FDTD.¹⁸ Several individual hyperthermia applicators have been simulated using FDTD.¹⁹ FDTD requires N^2 computations, where N is a cell in the (normally cubical) FDTD grid. Unlike method of moments, every cell in space (including at least a minimal amount of air surrounding the model) must be included in the discrete model, so the total number of cells, N , is likely to be larger. However, the significant improvement in computational efficiency generally makes this trade-off favor FDTD for bioelectromagnetic simulations. Complete detailed analysis of breast cancer imaging modalities was also done

with FDTD,^{20–43} as well as hyperthermia systems,⁴⁴ and evaluation of cell phones (including those with dual antennas) near the human head.^{45,46} Antennas for implantation in the body (mostly microstrip or Planar Inverted F (PIFA) types) have been simulated with FDTD and in some cases optimized with genetic algorithms.⁴⁷ Deep hyperthermia applicators (annular phased arrays) have been simulated extensively with FDTD.^{48,49}

Several FDTD developments have been important for bioelectromagnetic simulations, including the development of frequency-dependent methods (FD)²TD,⁵⁰ low-frequency FDTD methods,⁵¹ efficient FDTD computation,⁵² and evaluation of temperature using the bioheat equation.⁴⁹

Model development is one of the significant challenges of numerical bioelectromagnetics. Models have progressed from the prolate spheroidal models of the human used during the 1970s⁵³ to roughly 1-cm models based on anatomical cross sections used during the 1980s⁵⁴ to a new class of millimeter-resolution MRI-based models of the body that have been the hallmarks of research since the 1990s.^{55–58} Today, probably the most widely used models are derived from the Visible Man Project.⁵⁹

Once a tissue-segmented model has been chosen, the electrical properties of the tissues are defined. The properties of human tissue vary significantly with frequency, so it is essential to use data accurately measured at the frequency of interest. There is a wide range of published data on measured tissue properties,^{53,60–71} and work is still underway to measure and verify these properties. These and other references are electronically searchable at the University of Utah Dielectric Database OnLine.⁷²

38.3 ANTENNAS FOR MEDICAL IMAGING

One of the most promising uses of antennas in medical applications is for imaging the location of leukemia,⁷³ breast tumors,^{20–43} and cardiac anomalies.^{74,75} Microwave imaging methods rely on the fact that the electrical properties of normal and malignant tissue are significantly different^{60–67} and that there is significant variation from tissue to tissue. Location of breast cancer shows particular promise, because its relatively low loss allows electromagnetic fields to propagate to the tumor and back, and the proximity of the tumor to the outer surface of the body means that the signal does not have more than a few inches to propagate. Two major microwave imaging methods utilize antenna arrays. Tomography^{20–27} attempts to map a complete electrical profile of the breast, and confocal imaging^{28–43} maps only the location of significant scatterers. Both of these methods have used antenna arrays made up of wideband elements to send and receive the test signals. Microwave tomography picks up the passive electromagnetic fields from the body.^{76–92} Magnetic Resonance Imaging (MRI) uses a strong magnetic field to cause the magnetic dipoles in the body to precess and then uses an array of loops to pick up the fields when they relax back to their normal state.^{11,93–113}

Tomography for Breast Cancer Detection

Microwave tomography is used to provide a complete spatial mapping of the electrical properties in the region of interest. During the acquisition phase, an array of antennas surrounds the region of interest. One of the antennas in the array is used to transmit a signal, normally a sine wave,²⁰ set of sine waves,^{21–22} or a broadband signal,²³ and all of the other antennas are used to receive the reflected signal. The array is scanned so that each antenna transmits each frequency, and these signals are received by each of the other antennas. After all of the data has been acquired, it is processed by comparing the received data with what

would be expected from a simulated model of the region. A numerical “forward model” is used to predict how much power is transmitted from the transmit antenna, passes into and reflects from the breast/tumor model, and is received by the receive antenna. Originally, the simulated model is just a good guess for what might be present, generally a generic breast model with no tumor. The differences between the measured and expected received data are used to modify the original guess to obtain an ideal model that best matches the measured data. This “inversion” is used to predict what model could have produced the measured data.

Microwave tomography for breast cancer has been demonstrated by several groups.^{20–27} In the Dartmouth system,²⁰ for instance, sine waves from 300–1000 MHz (being expanded to 3 GHz) are transmitted from a circular array of 16 transmit/receive monopole antennas, shown in Figure 38-2, to produce 2D reconstructed images of the breast. Quarter-wave monopole antennas (in the fluid) that were built by extending the inner conductor of semi-rigid coax were used for this application. Monopoles were chosen because they are easy to model as a line source in a 2D reconstruction algorithm with high accuracy.²⁰ Water-filled waveguide apertures have also been used for tomography, but the monopole antennas were found to be as accurate and easier to build.²⁴

Microwave tomography has been validated experimentally.²⁷ The presence of 1.1- or 2.5-cm saline tubes (representing tumors) in excised breast tissue are seen to be clearly visible.²⁶ Objects as small as 4 mm in diameter have been imaged at 900 MHz.²⁷

Confocal Imaging for Breast Cancer Detection

Confocal imaging for breast cancer detection is another exciting application of antenna arrays in medical imaging. Confocal imaging is similar to ground-penetrating radar. Unlike microwave tomographic imaging, this method does not provide a complete electrical mapping of the region of interest. Instead it identifies locations of significant scattering. This method typically uses a single antenna scanned in a flat array pattern above the breast or a cylindrical array of very small broadband antennas.²⁸ For planar imaging, the patient lies face up, and the antenna is physically scanned in a plane above the breast.^{29–31} For cylindrical imaging, the patient lies face down, with the breast extending into the cylindrical array

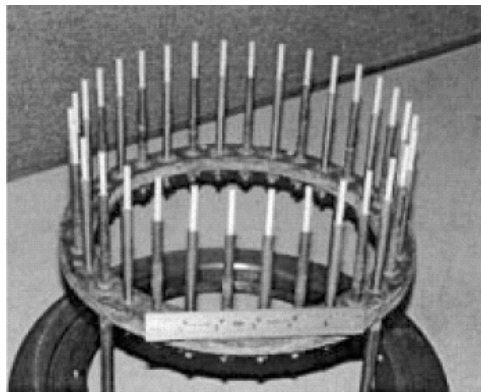


FIGURE 38-2 2D monopole array used for tomographic imaging of the breast (after P. M. Meaney et al²⁰ © IEEE 2000)

through a hole in the table.^{32,33} Matching fluid surrounding the breast, similar to that used for microwave tomography, is suggested in this case. Both methods provide similar results.³³ One antenna in the array transmits an ultrawideband (UWB) pulse, which propagates into the breast, where it is reflected off significant electrical discontinuities, and is received in parallel by the other antennas in the array. Knowing the physical spacing between the array elements, the different delays between the transmit antenna, scattering point, and receiving antenna can be calculated geometrically. The received pulses representing a specific point in space can then be time delayed appropriately for each antenna, added up, and integrated to indicate the magnitude of the scattered energy from that point in space. This is effectively correlating the signals received from that point at all antennas.

The antennas used for confocal imaging must be UWB and small enough to fit within the relatively small array area. Resolution of less than 1 cm requires a bandwidth of at least 5 GHz. The lossy nature of tissue attenuates high-frequency signals, limiting the upper frequency to about 10 GHz. Initially, resistively loaded bowties were suggested for the planar configuration,^{29–31,35,37} while dipole antennas were suggested for the cylindrical system.^{32,33} Resistively loaded Vee dipoles have also been proposed.³⁶ In the cylindrical configuration, multiple antennas are present in the array, although they are not simultaneously active. In the planar system, a single antenna is scanned over the surface, creating a synthetic antenna aperture. To overcome the inherent inefficiency of resistively loaded antennas, a modified ridged horn antenna operating from 1 to 11 GHz has been introduced.³⁸ Most of the antennas are designed to observe copolarized reflections from the breast; however, two resistively loaded bowtie antennas in the shape of a Maltese cross, as shown in Figure 38-3, have also been proposed to pick up the cross-polarized reflections.³⁰ Cross-polarized reflections from simple tumor models were also examined.^{34,39}

The antenna shown in Figure 38-3³⁴ consists of two cross-polarized bowtie antenna elements, an octagonal cavity behind the bowtie elements, and a metal flange attached to the cavity. The broadband bowties have flare angles of 45°. They are 1.67 cm long, which is a half-wavelength at 3 GHz in fat (similar to breast). The octagonal cavity blocks waves radiated away from the breast. The cavity is approximated as a circular waveguide filled with fat material for matching and size reduction. The first cutoff frequency is set to be

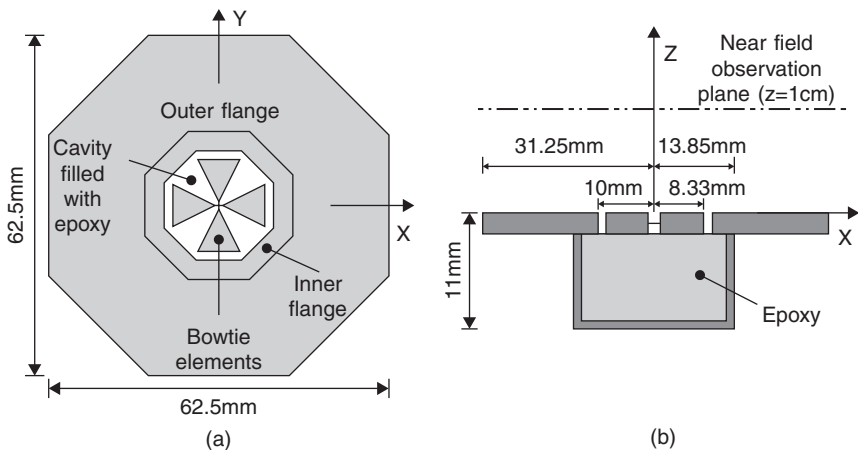


FIGURE 38-3 Cross-polarized antenna for confocal imaging. The properties of the substance inside the cavity and the medium outside the antenna are similar to fat ($\epsilon_r = 9$; $\sigma = 0.2$ S/m) (after X. Yun et al³⁴ © IEEE 2005).

2 GHz for 2 to 4 GHz operation. The cavity length is a quarter wavelength, which is 11 mm at 3 GHz. The flange consists of an inner and outer component, and is designed to block unwanted waves such as surface waves. The antenna performance does not change significantly when the flange size is varied between 10–6.25 mm, therefore, the width of the outer flange is set to be 6.25 mm. The inner flange is designed to prevent possible electric field overshoot at the inner corners of the opening of the octagonal cavity or at the ends of the bowtie elements. A slotline bowtie antenna has also been proposed.⁴⁰

A resistively loaded monopole antenna, shown in Figure 38-4, suitable for use in a cylindrical array was proposed by Sill and Fear.^{41–43} Based on the Wu–King design^{114–115} this antenna was designed to be usable from 1 to 10 GHz immersed in canola oil ($\epsilon_r = 3.0$) for matching to breast tissue. The antenna is fabricated using high-frequency chip resistors (Vishay 0603HF, by Vishay Intertechnology, Malvern, PA) soldered to a high-frequency substrate (Rogers RO3203 series by Rogers Corporation, Chandler, AZ). The substrate ($\epsilon_r = 3.02$ and $\sigma = 0.001$ S/m) has electrical properties similar to those of the canola oil. The antenna is soldered to a subminiature A (SMA) connector and attached to a metal plug for connection into the oil-filled test canister.

The cylindrical confocal imaging system has been experimentally tested.^{41,43,116} Simulated tumors as small as 4 mm have been detected using a 3D system.

Microwave Radiometry

Microwave radiometry is a passive method where the natural electromagnetic radiation or emission from the body is measured to allow detection or diagnosis of pathogenic conditions.⁷⁷ This method has been proposed for detection of breast cancer^{80,90,91} and brain cancer,⁸¹ in which the metabolism of cancer cells increases the localized temperature 1 to 3°C. This method has also been used for fluid and blood warming,⁷⁷ for detection of rheumatology,⁷⁹ and for monitoring temperature rise during hyperthermia treatment.⁷⁸

Typical antennas include open-ended rectangular waveguides,^{82–85} small-loop antennas,⁸⁷ and a horn antenna with a dielectric lens.⁸⁸ Working around 3 GHz, all of these antennas have radiation patterns that have minimal penetration into the body, thus strongly weighting them to monitoring of surface temperatures.^{89,92} Increased focus, and therefore better spatial accuracy, was obtained with an array of six rectangular aperture antennas filled with low-loss



FIGURE 38-4 Fabricated resistively loaded monopole antenna soldered to an SMA connector and attached to a metal plug (after J. M. Sill and E. C. Fear⁴³ © IEEE 2005)

dielectric ($\epsilon_r = 25$), which were scanned over the object of interest in an overlapping pattern. Preliminary results indicated promise for location of breast tumors.⁹⁰⁻⁹¹

Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) uses a very strong magnetic field (0.5T, 1.5T, 3T, 4T, 7T, perhaps in the future 8T) to make the magnetic dipoles in the body precess (line up). When they are released, a set of receiver coils picks up the magnetic field created when these dipoles return to their normal orientations (position may change a lot as in blood imaging, diffusion, etc.). The relaxation properties of the different tissues affect the relative received signal intensities, and a 3D map of the body can be produced.

There are two basic types of receiver coils used for MRI. Volume coils, such as the quadrature birdcage head coil shown in Figure 38-5,¹¹ are used for imaging large and deep anatomic structures of the body and provide homogeneous field profiles. For high-resolution applications that are more localized, such as angiographic imaging, hippocampus imaging, and functional imaging, in which the object features are very small, volume coils pick up less signal and more noise, thus having a lower signal-to-noise ratio (SNR) and poor-quality images. Modifications such as the use of an RF reflector or “endcap,”¹¹ and modified shapes such as the elliptical⁹⁴ or “dome”^{95,96} coils, have been developed. Smaller-volume surface coils⁹⁷ have been shown to improve image quality, particularly when combined into phased arrays⁹⁸⁻¹⁰⁶ such as the one shown in Figure 38-6. Phased-array coils are closer to the area of interest, so they pick up larger signal strength, and are smaller, so they pick up less noise, thus having a higher SNR. They are designed to overlap so that the mutual inductance between coils is zero, and so that the impedance at the preamplifier is very low, for optimal SNR.⁸ Part of the price for this improved image quality is the complexity of the receiver and data acquisition system, as each antenna is received on an individual channel. The image processing is also more computationally expensive, as the signal from each antenna is weighted depending on its proximity to the target region (and hence expected relative SNR), phase shifted, and combined with the other similarly processed signals.

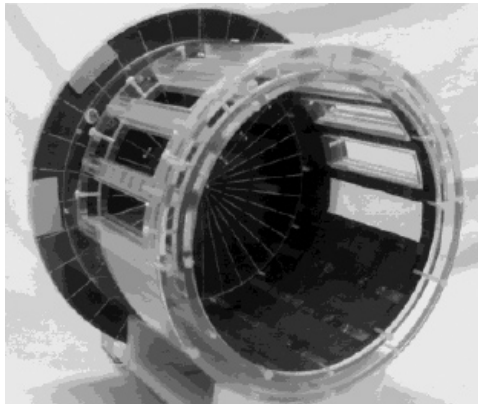


FIGURE 38-5 Quadrature birdcage coil with endcap used for whole-volume head imaging (reprinted with permission from J. R. Hadley et al¹¹ © *Journal of Magnetic Resonance Imaging* 2000)

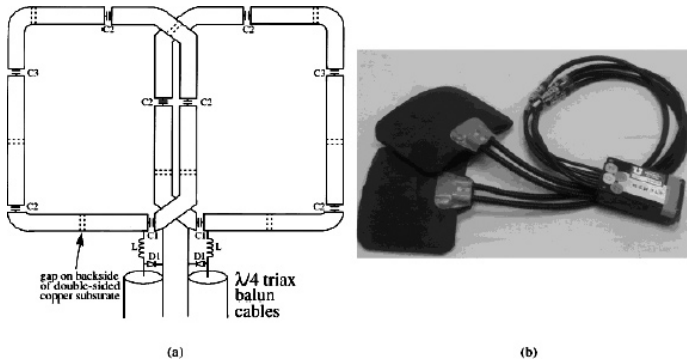


FIGURE 38-6 (a) Two-element phased-array coil design. Dashed lines indicate the breaks in the underside of the double-sided copper section of the coil. (b) Image of finished phased-array coils (enclosed in foam) with triax balun cables and phased-array port connector box (reprinted with permission from J. R. Hadley et al¹¹ © *Journal of Magnetic Resonance Imaging* 2000).

Among the practical considerations that are challenging with phased-array coils are the data acquisition time and the limited field of view, particularly for applications where the region of interest (an arterial occlusion, for instance) may not be precisely known and is therefore easy to miss. Phased-array coils have been used for numerous magnetic resonance angiography (MRA) applications including peripheral^{107,108} abdominal, intracranial, and carotid imaging.^{109–111} Recent coil designs have started to integrate phased-array elements into volume-like coils with the ability to control how the image is constructed to achieve maximum image quality.^{112,113} For these applications, the coil array functions much like the phased array in a synthetic aperture radar application. The image quality for the different coil types and configurations depends strongly on the application. The optimal image construction algorithm depends strongly on the application and region of interest, making the flexibility of being able to synthetically develop large or small subarrays very attractive.

38.4 HEATING

Hyperthermia

Hyperthermia (HT)^{117,118} is a method of treating cancer by heating the body. The tissue is typically heated to 41 to 45°C for 30 to 60 minutes. Often, this involves focusing the energy on the tumor region, relying on the tumor to be more sensitive to heat than the surrounding healthy tissues. This may be due to poor internal vasculature of the tumor or its higher conductivity and permittivity caused by increased water and ionic content. HT has also been shown to increase the effectiveness of radiation or chemotherapy.^{119–120} The most commonly used frequencies for hyperthermia are 433, 915, and 2450 MHz. The type of antenna or antenna array used for HT depends on whether it is to be administered superficially, interstitially, or deep-body.

Superficial HT applicators include microstrip,¹²¹ waveguide,¹²² current sheets,¹²³ inductive,¹²⁴ and the dual concentric conductor antenna, or DCC, shown in Figure 38-7.^{125–128} The DCC is particularly attractive, because it can be easily fabricated on flexible, printed,

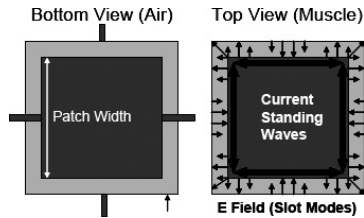


FIGURE 38-7 DCC antenna geometry, E fields in the ring slot and edge currents (after P. F. Maccarini et al.¹²⁶ © IEEE 2004)

circuit-board material, which makes it easy to conform to virtually any part of the human body. The DCC aperture is a ring-slot configuration fed simultaneously on all four sides. Prediction and optimization of the heating is normally done by analyzing the near fields of the antenna (the heating region) numerically.¹²⁶

Interstitial applicators for HT are typically monopole antennas made from coaxial cables with the center conductor extending beyond the outer ground shield of the cable.¹²⁹ These antennas have a tear-drop-shaped radiation pattern, so the majority of the heating is near the feedpoint of the antenna (where the ground shield stops), leaving the tip of the antenna extended beyond the useable heating range. The heating distribution can be made more uniform by varying the width of the conductor¹⁶ or adding a choke to the antenna.¹³⁰ Numerous other designs of interstitial applicators also exist. The heating pattern can be adjusted within the array by phasing the antenna elements¹⁶ or by using nonuniform insulation.¹³¹

Deep-body HT applicators are generally based on annular phased arrays (APA) of waveguides,¹¹⁸ coaxial TEM apertures,¹³² printed antennas,^{129,133} and induction systems.¹³⁴ Originally, APA systems contained only one ring of 2D applicators surrounding the patient. The ring could be scanned vertically. Significant improvement with a true 3D HT system with the applicators vertically offset has been observed.¹²⁹ The first clinically used 3D-type applicator is the SIGMA-Eye applicator (BSD Medical Corp., Salt Lake City, UT¹³⁵). A detailed description of this applicator and different numerical antenna feed models can be found in Nadobny et al.¹³⁶

Among the ongoing antenna design challenges in this area is the design of antennas that can be used to also monitor temperature and administer radiation therapy.^{128,133,137} One prototype combination device is shown in Figure 38-8 and another in Figure 38-9. Another research area is the use of optimization approaches to predict and control the heating pattern.^{138,139}

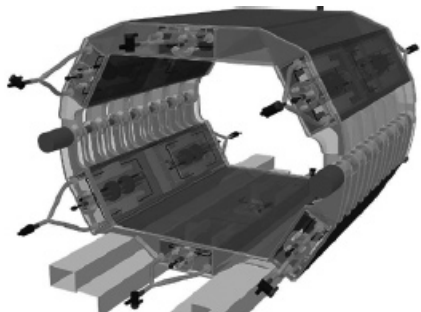


FIGURE 38-8 A prototype of the new Berlin MR-compatible Water-Coated Antenna Applicator (WACO) (after J. Nadobny et al.¹³⁷ © IEEE 2005)

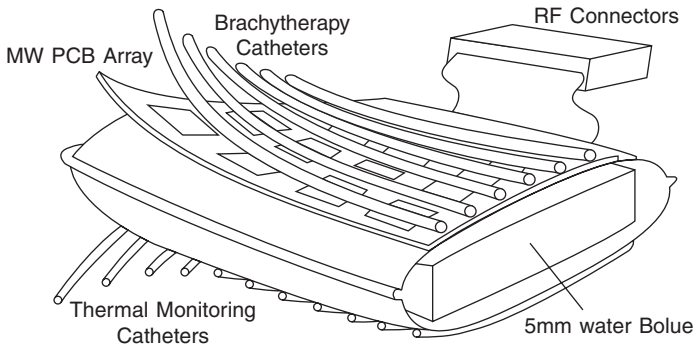


FIGURE 38-9 Schematic of combination applicator showing component parts: parallel catheter arrays for brachytherapy sources and thermal mapping sensors, PCB antenna array, and water coupling bolus (after P. Stauffer et al¹³³ © IEEE 2004)

Treatment of Hypothermia

Patients who have experienced hypothermia are at great risk during the rewarming process. Conventional methods such as warm water baths use external sources of heat, which warm the peripheral regions while the heart is still cold. The demand for increased circulation to the extremities can overload the heart. EM applicators can produce deeper heating than methods that simply heat the body surface and rely on thermal conduction to carry the heat to the deeper tissues. EM heating provides heat deeper into the body, which increases cardiac output and circulates warmed blood to the peripheral tissues without overloading the heart.¹⁴⁰

Cardiac Ablation

Microwave catheter ablation uses monopole antennas inserted in catheters to the heart to treat cardiac arrhythmias by creating deep lesions that destroy the source of the arrhythmias. Frequencies of 915 or 2450 MHz are typically used.^{141,142} Ablation requires localized temperatures of 50–90° C for a short time, typically a few minutes. Microwave ablation has also been used in conjunction with traditional balloon angioplasty to soften arterial plaque.¹⁴³

38.5 COMMUNICATION (BIOTELEMETRY)

Wireless communication systems and their associated implantable antennas are needed for communication with implantable medical devices such as cardiac pacemakers and defibrillators,¹⁴⁴ neural recording and stimulation devices,¹⁴⁵ and cochlear¹⁴⁶ and retinal¹⁴⁷ implants. Designing antennas for embedded applications is extremely challenging because of reduced antenna efficiency, impact of the environment on the antenna, the need to reduce antenna size, and the very strong effect of multipath losses. In addition to the present needs for embedded antennas, the expansion of MEMS, which are expected to play a dominant role in next-generation technologies, will add dramatically to the applications for imbedded antennas. Ultra-small devices (small enough to be injected in a human vein, for instance) and the desire to communicate with them will inevitably lead to the need for miniaturized antennas embedded in lossy environments. Emerging medical telemetry devices have led

to recent advances in the design of small, biocompatible antennas that can be implanted in the human body.^{148–175}

Major challenges exist for implantable antennas. Not only do they have to be long-term biocompatible, but they must also be small, passive or highly efficient (to reduce battery requirements inside the body), and able to transmit power through the highly lossy body structure. In addition, they must meet the maximum SAR guidelines, which can be quite challenging. The very near field of the antenna is the inductive zone where power is not radiated. If lossy material is in the inductive zone of the antenna, which it normally is for implanted antennas, this near field power is absorbed and shows up in the SAR measurements or calculations. Thus, SAR is often the limiting factor for power transmitted from an implanted device.

In the majority of wireless telemetry cases for implantable devices, inductively coupled coils often wound around a dielectric or ferrite core are used.^{152–167,176} In these cases, the time-varying magnetic field generated by the primary coil is received by the secondary coil, which results in an induced current in the implanted coil. Frequencies are often lower than 50 MHz to ensure that the presence of the human body does not significantly obstruct the coupling between the coils. In this case, the most important parameters for the design of the telemetry system are the self and mutual inductances of the coils. Several methods can be used to determine these parameters depending upon the frequency of interest and the geometrical shape of the coils. For the simplest traditional coil geometries (circle, square), analytical approximations of self and mutual inductances are often used, whereas more sophisticated methods (such as the partial inductance method and similar methods) can be used in the case of geometrically complex coils. A measure of the quality of coupling between two coils is given by the coupling coefficient K between two coils ($0 \leq K \leq 1$), defined by

$$K = \frac{M_{12}}{\sqrt{L_1 L_2}}$$

where M_{12} is the mutual coupling between coils 1 and 2, and L_i is the self inductance of coil i . To maximize the power delivered to the load in these applications, usually a capacitor is inserted in parallel with the inductance of the coil and the resistance of the load to form a parallel resonant LC circuit. Many other parameters may affect the design of inductively coupled coils for biomedical telemetry systems, such as implant size, maximum power, temperature increase in the implanted device, and specific absorption rate of power (SAR given in W/kg) induced in the tissue.¹⁷⁷

Most inductive telemetry links are used for subcutaneous applications due to power restrictions for passive devices. Data rates are generally low, and size/weight and biocompatibility issues plague these devices. However, recent advances continue to reduce the power requirements and provide more biocompatible designs. The Utah Electrode Array (UEA), for example, uses a pickup coil printed on a ceramic substrate and integrated with the implanted neural electrode array, as shown in Figure 38-10.¹⁴⁵ The implanted coil is energized by an external inductive programmer/reader that powers the implanted circuitry while transferring telemetry data.

Radiofrequency links are also being developed for communication with medical implants. For cardiac telemetry, a dipole¹⁷¹ and spiral or serpentine microstrip,^{173,174} and a waffle-type patch¹⁷⁸ have been designed for implantation in the shoulder. An insulated wire antenna has also been used, and this wire may be used as the lead between the heart and the battery pack/controls of the pacemaker.¹⁶⁸ The antenna can be treated as a waveguide, where the lossy body acts as the outer conductor of the waveguide. The insulated antenna in tissue may be matched with a load resistor connected to the conducting tissue in order to reduce or eliminate the reflection.¹⁶⁹

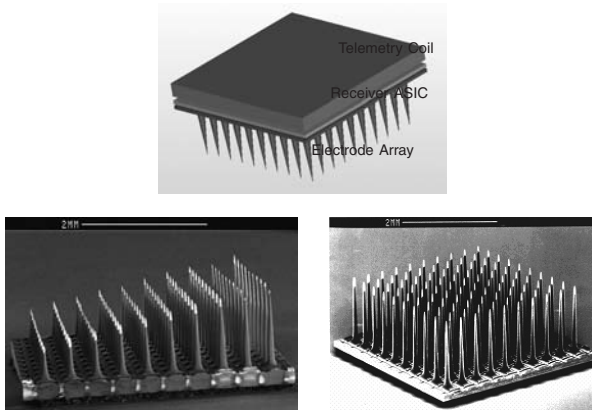


FIGURE 38-10 Utah Electrode Array packaged with a custom ASIC and printed receiver coil (after K. Guillory and R. A. Normann¹⁴⁵ © *J. Neurosci. Methods* 1999)

Another type of antenna used for communication with cardiac devices is the circumference antenna, which is a monopole antenna that is mounted around the edge of the pacemaker case, as shown in Figure 38-11.^{169,170} The 94-mm-long circumference antenna shown in Figure 38-11 was centered in a plastic insulator with a thickness of 10 mm. The bandwidth where the SWR of this antenna is less than 2 is 42 MHz, which is larger than the required MICS allocation of 3 MHz.

For smaller implants, a microstrip patch antenna has been successfully used for a retinal prosthesis,¹⁷⁵ and a small dipole has been designed for communication with a brain implant.¹⁷²

Deep-torso devices will experience more loss than subcutaneous devices. Furthermore, the location in the body controls the radiation pattern shape as well as magnitude. For example, the calculated radiation patterns for a small, multiturn loop antenna implanted in



FIGURE 38-11 Circumference antenna on a pacemaker model (after A. Johansson,¹⁶⁹ Figure 5-3)

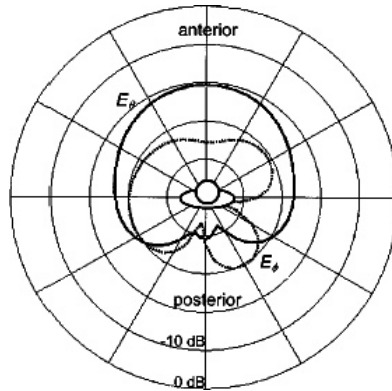


FIGURE 38-12 Calculated azimuthal radiation patterns for a 418-MHz vertical source: vertically polarized pattern (solid line) and horizontally polarized pattern (dotted line) (after W. G. Scanlon et al.¹⁵⁶ © IEEE 2000)

the vagina are shown in Figure 38-12. Measured net body losses (power absorbed in the body) for this 418-MHz antenna are 19.2 dB. The bodyworn radiation efficiency is

$$\eta_b = \frac{P_{\text{body}}}{P_{\text{air}}} = 1.2\%$$

where P_{air} is the total power radiated by the antenna in air, and P_{body} is the total power radiated (external to the body) by the antenna implanted in the body. At 916.5 MHz, the measured net body loss is 24.3 dB, and the bodyworn radiation efficiency is 0.37 percent.¹⁵⁶ The substantial losses in the body have so far limited deep-torso implants to communication with receivers held on or very near the body.

38.6 PULSED ELECTROMAGNETIC FIELDS

Pulsed electromagnetic fields (PEMFs) have been developed for a number of medical applications. These fields are generally delivered by electrodes connected directly to the body, and as such are not truly an application of antennas. However, since this method is showing significant promise for many different medical applications, and since UWB antennas are being used in many other areas, it is not unlikely that antenna design concepts could be applied to PEMF applicators in future applications.

Bone and Tissue Healing

Pulsed electromagnetic fields have been found to be highly effective for healing fractures and soft-tissue injuries, particularly those that do not respond to ordinary healing methods. As early as 1812, passing “electric fluids” through needles inserted in the fracture gap was found to stimulate bone healing, and by the mid-1800s, this DC current stimulation was

considered the method of choice for slow-healing fractures. Today's bone-healing PEMF systems typically use a 72-Hz single-pulse pattern. While the full biological mechanism is not minutely understood, it appears PEMFs of this type are able to retard the osteoclasts that destroy bone while increasing the rate of new bone formation.¹⁷⁹

Pain Control

Pulsed electromagnetic fields have also been used for controlling pain. Electrodes are placed strategically around the knee, shoulder, back, etc., and PEMF is applied. This has been found to provide both short- and long-term pain relief, although the exact reason is not fully understood. This method is currently used in both human and veterinary medicine.¹⁷⁹

Drug Delivery and Electrochemotherapy

Pulsed electromagnetic fields can be used for a wide variety of needle-less drug delivery applications. Iontophoresis is a method to electrically force drugs across a transdermal interface using a relatively small voltage (0.1–10 V) across the skin boundary. This method appears not to create structural changes in the cells or the skin, but rather just creates ion pathways that a conductive fluid (drug) will follow through pre-existing aqueous pathways. At present, a limited number of drugs can be delivered using this method.¹⁸⁰

Pulsed electromagnetic fields can also be used to treat cancer, using a new therapy called electrochemotherapy, which has been used for a variety of cutaneous tumors, including head and neck tumors, melanomas, superficial breast cancer lesions, etc. In this therapy, the resistance of malignant cells to penetration by certain chemotherapeutic agents is temporarily lowered by electroporation, which creates temporary pores (pathways) in the membranes of the malignant cells by the application of short DC pulses that generate electric fields of several kV/cm. Once the cells are porated, the chemotherapeutic agents can enter the malignant cells and destroy them. Electrochemotherapy not only can increase the efficacy of certain chemotherapeutic agents, but also can reduce side effects because malignant cells can be destroyed with much lower doses of chemotherapeutic agents than with conventional chemotherapy.^{180,181}

38.7 SENSING

In addition to receiving and transmitting power for communication or imaging and depositing power for heating applications, antennas can be used as sensors. Antennas are used as probes for dielectric properties and electric or magnetic fields.

Dielectric Measurement Probes

Measurement of the electrical properties of tissues has been done extensively to facilitate research, numerical modeling, etc. High-frequency in vivo and in vitro dielectric measurements of tissue are typically made using an open-ended coaxial probe.¹⁸² The coaxial probe is sensitive to material that lies within a fringing capacitance zone adjacent to the probe tip. A two-wire, dipole-type probe has also been used.¹⁸³ Another application of dielectric property measurements is in vivo measurement of brain fluid.¹⁸⁴

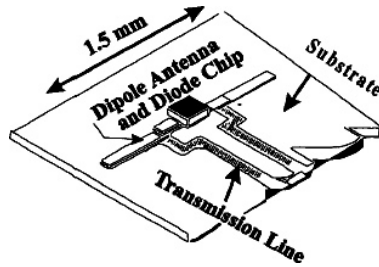


FIGURE 38-13 Miniature printed dipole antenna for measurement of electric fields to determine cell phone RF exposure compliance (H. Bassen and G. Smith¹⁸⁵ © IEEE 1983)

Electric and Magnetic Field Probes

Miniature electric field probe antennas have been designed for assessment of compliance of electromagnetic devices with RF exposure guidelines.¹⁸⁵ Measurement of SAR requires evaluation of the localized electric field, which should not be perturbed by the probe. This requires a very tiny electric or magnetic field receiver, such as the one shown in Figure 38-13. Because this probe is inherently sensitive to the polarization of the electric field, three perpendicular probes are used, as shown in the SPEAG probe in Figure 38-14. A magnetic field probe is also shown in this figure, with three perpendicular receiving loops.¹⁸⁶

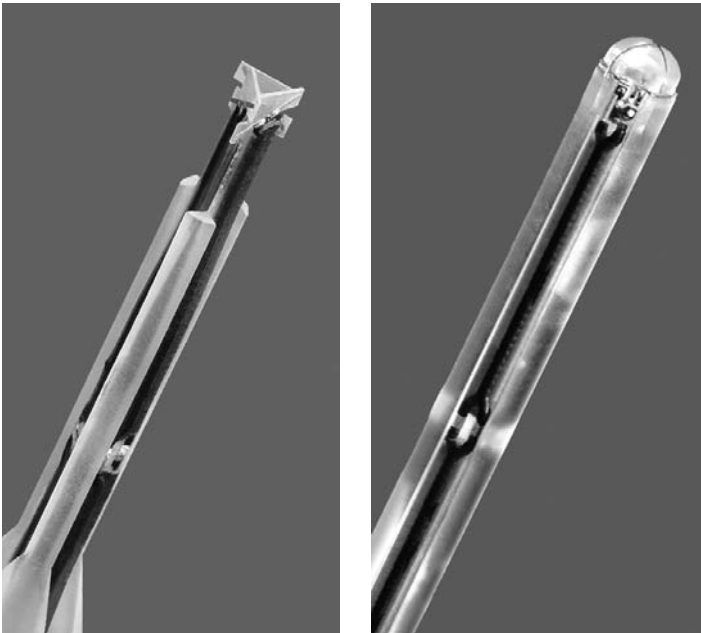


FIGURE 38-14 Electric and magnetic field probes from SPEAG (reprinted with permission, Schmid & Partner Engineering AG, Zurich)

38.8 FUTURE DIRECTIONS

The medical applications of antennas described in this chapter are by no means all inclusive. New technology is rapidly being developed, and creative new ideas continue to emerge. The basic capabilities of antennas to transmit information, deliver heat, sense electrical properties, and receive information for imaging will continue to lead to new applications for antennas in medicine. At the risk of tabloid-type predictions, the following are some expectations for where antennas will continue to grow in medical applications.

Communication with medical devices is an area that is rapidly expanding. The first medical devices were cardiac devices that had large battery packs and control systems and minimal data up/downlink requirements that could be managed in a doctor's office. Today these devices are pressing for higher data rates, real-time communication, and more efficient links. Their size has been continually shrinking and promises to shrink radically, due mainly to advances in battery technology. Emerging nerve stimulation or recording devices require far less power than cardiac devices, and therefore do not require batteries at all. Prosthetic nerve devices hold promise for artificial vision, hearing, smell, balance and muscle control, nerve "repair," and a new level of treatment of brain malfunction for Parkinson's disease, depression, epilepsy, and more. New packaging techniques and ultra-efficient electronics are driving the need for superminiaturized antennas. This demand is likely to grow dramatically with the success of microscale electronics, microfluidics, and microscale sensors and actuators. Smart pills have shown the capability of "untethered" communication systems in the body, and other applications are likely to utilize this freedom in the future, such as moving through the bloodstream for diagnosis or treatment.

Devices used for heating have either been large, external devices that attempt to focus power, or much smaller devices inserted through a vein/artery or laproscopic surgical opening. These devices are also likely to shrink in size with the miniaturization of electronics, providing opportunity for more precise control of heat delivery.

Although medical imaging is a relatively mature field, new methods and radical enhancements to mainstream methods continue to emerge. New antenna designs for imaging tend to be larger numbers (arrays) of small antennas. Enhancing the bandwidth while miniaturizing the antenna continues to be a focus in many applications. Arrays that are nonuniform and nonplanar are likely to be important in many applications. With the ability to place antennas in the body for medical implant devices, it would not be surprising to see the development of imaging systems that can be swallowed, injected, or placed in body orifices for better imaging of sensitive structures.

Antennas are an integral part of medical devices today, and hold promise to play a significant role in the development of emerging devices for future medical systems.

Acknowledgments

Several researchers have contributed to this chapter, including Dr. Gianluca Lazzi (inductive coil telemetry), Jeff Johnson and Dr. Pichitpong Soontornpipit (biomedical telemetry), and Dr. Rock Hadley (MRI).

REFERENCES

1. Medical Implant Communications Service (MICS) Federal Register, "Rules and Regulations," vol. 64, no. 240 (Dec. 1999): 69926–69934.
2. International Telecommunication Union, Recommendation ITU-R SA.1346, 1998.

3. "Planning for Medical Implant Communications Systems (MICS) and Related Devices," Proposals Paper SPP 6/03, Australian Communications Authority.
4. M. H. Repacholi, *Biological Effects and Medical Applications of Electromagnetic Energy*, Chap. 2, O. P. Gandhi (ed.) (Englewood Cliffs, NJ: Prentice-Hall, 1990).
5. *IEEE Standard for Safety Levels With Respect to Human Exposure to Radiofrequency Electromagnetic Fields, 3 kHz to 300 GHz*, IEEE Std. C95.1, 1999, revised to *Standard for Safety Levels with Respect to Human Exposure to Radiofrequency Electromagnetic Fields, 3 kHz to 300 GHz*, IEEE Standards Coordinating Committee 28.4, 2006.
6. ICNIRP, "Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic, and Electromagnetic Fields (up to 300 GHz)," *Health Phys.*, vol. 74 (1998): 494–522.
7. National Institute of Environmental Health Science, <http://www.niehs.nih.gov/emfrapid/home.htm>.
8. O. P. Gandhi, *Biological Effects and Medical Applications of Electromagnetic Energy*, Chap. 3, O. P. Gandhi (ed.) (Englewood Cliffs, NJ: Prentice-Hall, 1990).
9. C. Gabriel, "Compilation of the Dielectric Properties of Body Tissues at RF and Microwave Frequencies," Final Technical Report, Occupational and Environmental Health Directorate Radiofrequency Radiation Division, Brooks Air Force Base, TX, 1996.
10. O. P. Gandhi, *Biological Effects and Medical Applications of Electromagnetic Energy*, Chap. 6 (Englewood Cliffs, NJ: Prentice-Hall, 1990).
11. J. R. Hadley, B. E. Chapman, J. A. Roberts, D. C. Chapman, K. C. Goodrich, H. R. Buswell, A. L. Alexander, J. S. Tsuruda, and D. L. Parker, "A Three-Coil Comparison for MR Angiography," *Journal of Magnetic Resonance Imaging*, vol. 11 (2000): 458–468.
12. R. W. P. King, G. J. Fikioris, and R. B. Mack, *Cylindrical Antennas and Arrays* (Cambridge, UK: Cambridge Univ. Press, 2002).
13. H. Massoudi, C. H. Durney, and M. F. Iskander, "Limitations of the Cubical Block Model of Man in Calculating SAR Distribution," *IEEE Trans. Microwave Theory and Tech.*, vol. 32 (1984): 746–752.
14. C. T. Tsai, H. Massoudi, C. H. Durney, and M. F. Iskander, "A Procedure for Calculating Fields Inside Arbitrarily-Shaped, Inhomogeneous Dielectric Bodies Using Linear Basis Functions with the Moment Method," *IEEE Trans. Microwave Theory and Tech.*, vol. 34 (1986): 1131–1139.
15. J. Rock Hadley, "Design of Radio Frequency Coil Arrays for Optimal Signal to Noise Ratio for Magnetic Resonance Angiography," PhD dissertation, University of Utah Electrical and Computer Engineering Department, 2005.
16. C. M. Furse and M. F. Iskander, "Three-dimensional Electromagnetic Power Deposition in Tumors Using Interstitial Antenna Arrays," *IEEE Trans. on Biomedical Engineering*, vol. 36 (Oct. 1989): 977–986.
17. O. H. Schaubert, D. R. Wilton, and A. W. Glisson, "A Tetrahedral Modeling Method of Electromagnetic Scattering by Arbitrarily Shaped Inhomogeneous Objects," *IEEE. Trans. Antennas and Propagation*, vol. 32 (1984): 75–82.
18. P. Cherry and M. F. Iskander, "FDTD Analysis of Power Deposition Patterns of an Array of Interstitial Antennas for Use in Microwave Hyperthermia," *IEEE Trans. Microwave Theory and Tech.*, vol. 40, no. 8 (Aug. 1992): 1692–1700.
19. J. C. Camart, D. Despretz, M. Chive, and J. Pribetich, "Modeling of Various Kinds of Applicators Used for Microwave Hyperthermia Based on the FDTD Method," *IEEE Trans. Microwave Theory and Tech.*, vol. 44, no. 10 (Oct. 1996): 1811–1818.
20. P. M. Meaney, M. W. Fanning, D. Li, S. P. Poplack, and K. D. Paulsen, "A Clinical Prototype for Active Microwave Imaging of the Breast," *IEEE Trans. Microwave Theory Tech.*, vol. 48 (Nov. 2000): 1841–1853.
21. W. C. Chew and J. H. Lin, "A Frequency-Hopping Approach for Microwave Imaging of Large Inhomogeneous Bodies," *IEEE Microwave Guided Wave Lett.*, vol. 5 (Dec. 1995): 439–441.
22. O. S. Haddadin and E. S. Ebbini, "Imaging Strongly Scattering Media Using a Multiple Frequency Distorted Born Iterative Method," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 45 (Nov. 1998): 1485–1496.

23. Q. Fang, P. M. Meaney, and K. D. Paulsen, "Microwave Image Reconstruction of Tissue Property Dispersion Characteristics Utilizing Multiple-Frequency Information," *IEEE Trans. Microwave Theory Tech.*, vol. 52 (Aug. 2004): 1866–1875.
24. P. M. Meaney, K. D. Paulsen, A. Hartov, and R. C. Crane, "An Active Microwave Imaging System for Reconstruction of 2-D Electrical Property Distributions," *IEEE Trans. Biomed. Imag.*, vol. 42 (Oct. 1995): 1017–1026.
25. K. D. Paulsen and P. M. Meaney, "Compensation for Nonactive Array Element Effects in a Microwave Imaging System: Part I—Forward Solution vs. Measured Data Comparison," *IEEE Trans. Med. Imag.*, vol. 18 (June 1999): 496–507.
26. P. M. Meaney, K. D. Paulsen, M. W. Fanning, and A. Hartov, "Nonactive Antenna Compensation for Fixed-Array Microwave Imaging: Part II—Imaging Results," *IEEE Trans. Med. Imag.*, vol. 18 (June 1999): 508–518.
27. P. M. Meaney, K. D. Paulsen, A. Hartov, and R. K. Crane, "Microwave Imaging for Tissue Assessment: Initial Evaluation in Multitarget Tissue-Equivalent Phantoms," *IEEE Trans. Biomed. Eng.*, vol. 43 (Sept. 1996): 878–890.
28. E. C. Fear, S. C. Hagness, P. M. Meaney, M. Okieniewski, and M. Stuchly, "Enhancing Breast Cancer Detection Using Near Field Imaging," *IEEE Microwave Magazine* (March 2002): 48–56.
29. S. C. Hagness, A. Taflove, and J. E. Bridges, "Two-dimensional FDTD Analysis of a Pulsed Microwave Confocal System for Breast Cancer Detection: Fixed-Focus and Antenna-Array Sensors," *IEEE Trans. Biomed. Eng.*, vol. 45 (Dec. 1998): 1470–1479.
30. S. C. Hagness, A. Taflove, and J. E. Bridges, "Three-dimensional FDTD Analysis of a Pulsed Microwave Confocal System for Breast Cancer Detection: Design of an Antenna-Array Element," *IEEE Trans. Antennas Propagat.*, vol. 47 (May 1999): 783–791.
31. X. Li and S. C. Hagness, "A Confocal Microwave Imaging Algorithm for Breast Cancer Detection," *IEEE Microwave Wireless Comp. Lett.*, vol. 11 (March 2001): 130–132.
32. E. Fear and M. Stuchly, "Microwave System for Breast Tumor Detection," *IEEE Microwave Guided Wave Lett.*, vol. 9 (Nov. 1999): 470–472.
33. E. C. Fear and M. A. Stuchly, "Microwave Detection of Breast Cancer," *IEEE Trans. Microwave Theory Tech.*, vol. 48 (Nov. 2000): 1854–1863.
34. X. Yun, E. C. Fear, and R. H. Johnston, "Compact Antenna for Radar-Based Breast Cancer Detection," *IEEE Trans. Antennas and Propagation*, vol. 53, no. 8 (Aug. 2005): 2374–2380.
35. S. C. Hagness, A. Taflove, and J. E. Bridges, "Wideband Ultralow Reverberation Antenna for Biological Sensing," *Electron. Lett.*, vol. 33, no. 19 (Sept. 1997): 1594–1595.
36. M. A. Hernandez-Lopez, M. Pantoja, M. Fernandez, S. Garcia, A. Bretones, R. Martin, and R. Gomez, "Design of an Ultra-broadband V Antenna for Microwave Detection of Breast Tumors," *Microw. Opt. Tech. Lett.*, vol. 34, no. 3 (Aug. 2002): 164–166.
37. E. C. Fear and M. A. Stuchly, "Microwave Breast Tumor Detection: Antenna Design and Characterization," *IEEE Antennas Propag. Symp. Dig.*, vol. 2 (2000): 1076–1079.
38. X. Li, S. C. Hagness, M. K. Choi, and D. W. W. Choi, "Numerical and Experimental Investigation of an Ultrawideband Ridged Pyramidal Horn Antenna with Curved Launching Plane for Pulse Radiation," *IEEE Antennas Wireless Propag. Lett.*, vol. 2 (2003): 259–262.
39. X. Yun, E. C. Fear, and R. H. Johnston, "Radar-based Microwave Imaging for Breast Cancer Detection: Tumor Sensing with Cross-Polarized Reflections," *IEEE Antennas Propag. Society Symp. Dig.*, vol. 3 (2004): 2432–2435.
40. C. J. Shannon, E. C. Fear, and M. Okoniewski, "Dielectric-Filled Slotline Bowtie Antenna for Breast Cancer Detection," *Electronics Letters*, vol. 41, no. 7 (March 2001).
41. J. M. Sill and E. C. Fear, "Tissue Sensing Adaptive Radar for Breast Cancer Detection: A Study of Immersion Liquid," *Electron. Lett.*, vol. 41, no. 3 (Feb. 2005): 113–115.
42. J. M. Sill and E. C. Fear, "Tissue Sensing Adaptive Radar for Breast Cancer Detection: Preliminary Experimental Results," *IEEE MTT-S Int. Microwave Symp. Dig.*, Long Beach, CA, June 2005.
43. J. M. Sill and E. C. Fear, "Tissue Sensing Adaptive Radar for Breast Cancer Detection—Experimental Investigation of Simple Tumor Models," *IEEE Trans. Microwave Theory Tech.*, vol. 53, no. 11 (Nov. 2005): 3312–3319.

44. D. M. Sullivan, "Three-dimensional Computer Simulation in Deep Regional Hyperthermia Using the FDTD Method," *IEEE Trans. Microwave Theory and Tech.*, vol. 38, no. 2 (Feb. 1990): 201–211.
45. B. M. Green and M. A. Jensen, "Diversity Performance of Dual-Antenna Handsets Near Operator Tissue," *IEEE Trans. Antennas and Propagation*, vol. 48, no. 7 (July 2000): 1017–1024.
46. O. Gandhi, G. Lazzi, and C. Furse, "Electromagnetic Absorption in the Human Head and Neck for Mobile Telephones at 835 and 1900 MHz," *IEEE Trans. on Microwave Theory and Techn.*, vol. 44 (1996): 1884–1897.
47. P. Soontornpipit, C. M. Furse, and Y. C. Chung, "Design of Implantable Microstrip Antenna for Communication with Medical Implants," Special Issue of *IEEE Trans. on Microwave Theory and Techn.* on Medical Applications and Biological Effects of RF/Microwaves (Sept. 2004).
48. C. E. Reuter, A. Taflove, V. Sathiaselan, M. Picket-May, and B. B. Mittral, "Unexpected Physical Phenomena Indicated by FDTD Modeling of the Sigma-60 Deep Hyperthermia Applicator," *IEEE Trans. Microwave Theory and Tech.*, vol. 46, no. 4 (April 1998): 313–319.
49. D. Sullivan, D. Buechler, and F. A. Gibbs, "Comparison of Measured and Simulated Data in an Annular Phased Array Using an Inhomogeneous Phantom," *IEEE Trans. Microwave Theory and Tech.*, vol. 40, no. 3 (March 1992): 600–604.
50. C. M. Furse, J.-Y. Chen, and O. P. Gandhi, "Use of the Frequency-Dependent Finite-Difference Time-Domain Method for Induced Current and SAR Calculations for a Heterogeneous Model of the Human Body," *IEEE Transactions on Electromagnetic Compatibility* (May 1994): 128–133.
51. C. Furse and O. P. Gandhi, "Calculation of Electric Fields and Currents Induced in a Millimeter-Resolution Human Model at 60 Hz Using the FDTD Method," *Bioelectromagnetics*, vol. 19, no. 5 (1998): 293–299.
52. C. M. Furse and O. P. Gandhi, "A Memory Efficient Method of Computing Specific Absorption Rate in CW FDTD Simulations," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 5 (May 1996): 558–560.
53. C. H. Durney, C. C. Johnson, P. W. Barber, H. Massoudi, M. F. Iskander, J. L. Lords, D. K. Ryser, S. J. Allen, and J. C. Mitchell, *Radiofrequency Radiation Dosimetry Handbook*, 2nd Ed. (Brooks AFB, TX: USAF School of Medicine, 1978).
54. O. P. Gandhi, Y. G. Gu, J. Y. Chen, and H. I. Bassen, "Specific Absorption Rates and Induced Current Distributions in an Anatomically Based Human Model for Plane-Wave Exposures," *Health Physics*, vol. 63, no. 3 (1992): 281–290.
55. O. P. Gandhi and C. M. Furse, "Millimeter-Resolution MRI-based Models of the Human Body for Electromagnetic Dosimetry from ELF to Microwave Frequencies," *Voxel Phantom Development: Proc. of an International Workshop*, Peter J. Dimbylow (ed.), National Radiological Protection Board, Chilton, UK, July 6–7, 1995.
56. P. J. Dimbylow, "The Development of Realistic Voxel Phantoms for Electromagnetic Field Dosimetry," *Voxel Phantom Development: Proc. of an International Workshop*, Peter J. Dimbylow (ed.), National Radiological Protection Board, Chilton, UK, July 6–7, 1995.
57. P. Olley and P. S. Excell, "Classification of a High Resolution Voxel Image of a Human Head," *Voxel Phantom Development: Proc. of an International Workshop* held at the National Radiological Protection Board, Chilton, UK, July 6–7, 1995, Peter J. Dimbylow (ed.).
58. M. A. Stuchly, K. Caputa, A. van Wensen, and A. El-Sayed, "Models of Human and Animal Bodies in Electromagnetics," *Voxel Phantom Development: Proc. of an International Workshop*, Peter J. Dimbylow (ed.), National Radiological Protection Board, Chilton, UK, July 6–7, 1995.
59. National Library of Medicine, Visible Man Project, MRI scans, CT scans, and photographs available on CD-ROM through Research Systems, Inc., 2995 Wilderness Place, Boulder, CO 80301.
60. C. Gabriel, S. Gabriel, and E. Corthout, "The Dielectric Properties of Biological Tissues: I. Literature survey," *Phys. Med. Biol.*, vol. 41 (1996): 2231–2249.
61. S. Gabriel, R. W. Lau, and C. Gabriel, "The Dielectric Properties of Biological Tissues: II. Measurements on the Frequency Range 10 Hz to 20 GHz," *Phys. Med. Biol.*, vol. 41 (1996): 2251–2269.

62. S. Gabriel, R. W. Lau, and C. Gabriel, "The Dielectric Properties of Biological Tissues: III. Parametric Models for the Dielectric Spectrum of Tissues," *Phys. Med. Biol.*, vol. 41 (1996): 2271–2293.
63. K. R. Foster and H. P. Schwan, "Dielectric Properties of Tissues and Biological Materials: A Critical Review," *Crit. Rev. Biomed. Eng.*, vol. 17 (1989): 25–104.
64. S. S. Chaudhary, R. K. Mishra, A. Swarup, and J. M. Thomas, "Dielectric Properties of Normal and Malignant Human Breast Tissues at Radiowave and Microwave Frequencies," *Indian J. Biochem. Biophys.*, vol. 21 (1984): 76–79.
65. A. J. Surowiec, S. S. Stuchly, J. R. Barr, and A. Swarup, "Dielectric Properties of Breast Carcinoma and the Surrounding Tissues," *IEEE Trans. Biomed. Eng.*, vol. 35 (April 1988): 257–263.
66. W. T. Joines, Y. Z. Dhenxing, and R. L. Jirtle, "The Measured Electrical Properties of Normal and Malignant Human Tissues from 50 to 900 MHz," *Med. Phys.*, vol. 21 (1994): 547–550.
67. A. M. Campbell and D. V. Land, "Dielectric Properties of Female Human Breast Tissue Measured in Vitro at 3.2 GHz," *Phys. Med. Biol.*, vol. 37 (1992): 193–210.
68. C. Gabriel, "Compilation of the Dielectric Properties of Body Tissues at RF and Microwave Frequencies," Final Report AL/OE-TR-1996-0037 submitted to Occupational and Environmental Health Directorate, RFR Division, 2503 Gillingham Dr., Brooks AFB, TX, June 1996.
69. M. A. Stuchly and S. S. Stuchly, "Dielectric Properties of Biological Substances – Tabulated," *J. Microwave Power*, vol. 15, no. 1 (1980): 19–26.
70. S. Rush, J. A. Abildskov, and R. McFee, "Resistivity of Body Tissues at Low Frequencies," *Circ. Research*, vol. XII (1963): 40–50.
71. L. A. Geddes and L. E. Baker, "The Specific Resistance of Biological Material—A Compendium of Data for the Biomedical Engineer and Physiologist," *Med. & Biol. Eng.*, vol. 5 (1967): 271–293.
72. University of Utah Dielectric Database OnLine, <http://www.ece.utah.edu/dielectric/>.
73. D. Colton and P. Monk, "A New Approach to Detecting Leukemia: Using Computational Electromagnetics," *IEEE Trans. Comput. Sci. Eng.*, vol. 2 (Winter 1995): 46–52.
74. S. Y. Semenov, A. E. Bulyshev, A. E. Souvorov, R. H. Svenson, Y. E. Sizov, V. Y. Borisov, V. G. Posukh, I. M. Kozlov, A. G. Nazarov, and G. P. Tatsis, "Microwave Tomography: Theoretical and Experimental Investigation of the Iteration Reconstruction Algorithm," *IEEE Trans. Micr. Theory Tech.*, vol. 46 (Feb. 1998): 133–141.
75. S. Y. Semenov, R. H. Svenson, A. E. Bulyshev, A. E. Souvorov, A. G. Nazarov, Y. E. Sizov, V. G. Posukh, and A. Pavlovsky, "Three-Dimensional Microwave Tomography: Initial Experimental Imaging of Animals," *IEEE Trans. Biomed. Eng.*, vol. 49 (Jan. 2002): 55–63.
76. K. L. Carr, "Microwave Radiometry: Its Importance to the Detection of Cancer," *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 12 (Dec. 1989): 1862–1869.
77. K. L. Carr, "Radiometric Sensing," *IEEE Potentials* (April/May 1997): 21–25.
78. L. Dubois, J.-P. Sozanski, V. Tessier, J.-C. Camart, J.-J. Fabre, J. Pribetich, and M. Chiv, "Temperature Control and Thermal Dosimetry by Microwave Radiometry in Hyperthermia," *IEEE Trans. Microwave Theory Tech.*, vol. 44, no. 10 (Oct. 1996): 1755–1761.
79. S. M. Fraser, D. V. Land, and R. D. Sturrock, "Microwave Thermography—an Index of Inflammatory Disease," *Br. J. Rheumatology*, vol. 26 (1987): 37–39.
80. B. Bocquet, J. C. Van de Velde, A. Mamouni, and Y. Leroy, "Microwave Radiometric Imaging at 3 GHz for the Exploration of Breast Tumours," *IEEE Trans. Microwave Theory Tech.*, vol. 38 (1990): 791–793.
81. J. Robert, J. Edrich, P. Thouvenot, M. Gautherie, and J. M. Escanye, "Millimeter Wave Thermography: Preliminary Clinical Finding on Head and Neck Diseases," *J. Microwave Power*, vol. 14 (1979).
82. K. L. Carr, A. M. El Mahdi, and J. Schaeffer, "Dual Mode Microwave System to Enhance Early Detection of Cancer," *IEEE Trans. Microwave Theory Tech.*, vol. 29 (1980): 256–260.
83. E. A. Cheever, J. B. Leonard, and K. R. Foster, "Depth of Penetration of Fields from Rectangular Apertures into Lossy Media," *IEEE Trans. Microwave Theory Tech.*, vol. 35 (1987): 865–867.

84. J. Audet, J. C. Bolomey, C. Pichot, D. D. n'Guyen, M. Robillard, M. Chive, and Y. Leroy, "Electrical Characteristics of Waveguide Applicators for Medical Applications," *J. Microwave Power*, vol. 15 (1980): 177–186.
85. A. W. Guy, "Electromagnetic Fields and Relative Heating Patterns Due to a Rectangular Aperture Source in Direct Contact with Bilayered Biological Tissue," *IEEE Trans. Microwave Theory Tech.*, vol. 29 (1971): 214–223.
86. D. V. Land, "Medical Microwave Radiometry and Its Clinical Applications," *IEE Colloquium Application of Microwaves in Medicine* (28 Feb. 1995): 2/1–2/5.
87. B. Enander and G. Larson, "Microwave Radiometry Measurements of the Temperature Inside a Body," *Electronic Letters*, vol. 10 (1974): 317.
88. J. Edrich and P. C. Hardee, "Thermography at Millimeter Wavelengths," *Proc. IEEE*, vol. 62 (1974): 1391–1392.
89. E. A. Cheever and K. R. Foster, "Microwave Radiometry in Living Tissue: What Does It Measure?" *IEEE Trans. Biomedical Engineering*, vol. 39, no. 6 (June 1992): 563–568.
90. B. Bocquet, J. C. van de Velde, A. Mamouni, Y. Leroy, G. Giauza, J. Delannoy, and D. Delvallee, "Microwave Radiometric Imaging at 3 GHz for the Exploration of Breast Tumors," *IEEE Trans. Microwave Theory Tech.*, vol. 38, no. 6 (June 1990): 791–793.
91. L. Enel, Y. Leroy, J. C. Van de Velde, and A. Mamouni, "Improved Recognition of Thermal Structures by Microwave Radiometry," *Electron. Lett.*, vol. 20 (1984): 293–294.
92. Y. Leroy, A. Mamouni, J. C. Van de Velde, B. Bocquet, and B. Dujardin, "Microwave Radiometry for Noninvasive Thermometry," *Automedica* (special issue on noninvasive thermometry), vol. 8 (1987): 181–201.
93. C. E. Hayes, W. A. Edelstein, J. F. Schenck, et al, "An Efficient, Highly Homogeneous Radiofrequency Coil for Whole-body NMR Imaging at 1.5 T," *J. Magn. Reson. Imaging*, vol. 63 (1985): 622–628.
94. M. C. Leifer, "Theory of the Quadrature Elliptical Birdcage Coil," *Magn. Reson. Med.*, vol. 38 (1997): 726–732.
95. S. Li, C. M. Collins, B. J. Dardzinski, et al, "A Method to Create an Optimum Current Distribution and Homogeneous B1 Field for Elliptical Birdcage Coils," *Magn. Reson. Med.*, vol. 37 (1997): 600–608.
96. J. R. Fitzsimmons, J. C. Scott, D. M. Peterson, et al, "Integrated RF Coil with Stabilization for FMRI Human Cortex," *Magn. Reson. Med.*, vol. 38 (1997): 15–18.
97. L. E. Hendrix, J. A. Strandt, D. L. Daniels, et al, "Three-Dimensional Time-of-Flight MR Angiography with a Surface Coil: Evaluation in 12 Subjects," *American Journal Radiology*, vol. 159 (1992): 103–106.
98. P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller, "The NMR Phased Array," *Magn. Reson. Med.*, vol. 16, no. 2 (1996): 192–225.
99. C. E. Hayes, N. Hattes, and P. B. Roemer, "Volume Imaging with MR Phased Arrays," *Magn. Reson. Med.*, vol. 18, no. 2 (1991): 309–319.
100. C. E. Hayes and P. B. Roemer, "Noise Correlations in Data Simultaneously Acquired from Multiple Surface Coil Arrays," *Magn. Reson. Med.*, vol. 16, no. 2 (1991): 181–191.
101. S. M. Wright, R. L. Magin, and J. R. Kelton, "Arrays of Mutually Coupled Receiver Coils: Theory and Application," *Magn. Reson. Med.*, vol. 17, no. 1 (1991): 252–268.
102. S. M. Wright and L. L. Wald, "Theory and Application of Array Coils in MR Spectroscopy," *NMR Biomed*, vol. 10, no. 8 (1997): 394–410.
103. G. R. Duensing, H. R. Brooker, and J. R. Fitzsimmons, "Maximizing Signal-to-Noise Ratio in the Presence of Coil Coupling," *J. Magn. Reson. B.*, vol. 111, no. 3 (1996): 230–235.
104. D. K. Sodickson and W. J. Manning, "Simultaneous Acquisition of Spatial Harmonics (SMASH): Ultra-fast Imaging with Radiofrequency Coil Arrays," *Magn. Reson. Med.*, vol. 38 (1997): 591–603.
105. K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity Encoding for Fast MRI," *Magn. Reson. Med.*, vol. 42 (1999): 952–962.
106. Y. Zhu, "Parallel Excitation with an Array of Transmit Coils," *Magn. Reson. Med.*, vol. 51, no. 4 (2004): 775–784.

107. K. Y. Kojima, J. Szumowski, R. C. Sheley, et al, "Lower Extremities: MR Angiography with a Unilateral Telescopic Phased-Array Coil," *Radiology*, vol. 196 (1995): 871–875.
108. J. W. Monroe, P. Schmalbrock, and D. G. Spigos, "Phased Array Coils for Upper Extremity MRA," *Magn. Reson. Med.*, vol. 33 (1995): 224–229.
109. C. E. Hayes, C. M. Mathis, and C. Yuan, "Surface Coil Phased Arrays for High-Resolution Imaging of the Carotid Arteries," *J. Magn. Reson. Imaging*, vol. 1 (1996): 109–112.
110. C. Yuan, J. W. Murakami, C. E. Hayes, et al, "Phased-Array Magnetic Resonance Imaging of the Carotid Artery Bifurcation: Preliminary Results in Healthy Volunteers and a Patient with Atherosclerotic disease," *J. Magn. Reson. Imaging*, vol. 5 (1995): 561–565.
111. S. H. Faro, S. Vinitski, H. V. Ortega, et al, "Carotid Magnetic Resonance Angiography: Improved Image Quality with Dual 3-inch Surface Coils," *Neuroradiology*, vol. 38 (1996): 403–408.
112. H. A. Stark and E. M. Haacke, "Helmet and Cylindrical Shaped CP Array Coils for Brain Imaging: a Comparison of Signal-to-Noise Characteristics," *Proceedings of the International Society for Magnetic Resonance in Medicine* (1996): 1412.
113. J. R. Porter, S. M. Wright, and A. Reykowski, "A 16-Element Phased-Array Head Coil," *Magn. Reson. Med.*, vol. 40 (1998): 272–279.
114. T. Wu and R. King, "The Cylindrical Antenna with Nonreflecting Resistive Loading," *IEEE Trans. Antennas Propag.*, vol. AP-13, no. 3 (May 1965): 369–373.
115. T. Wu and R. King, "Corrections to 'The Cylindrical Antenna with Nonreflecting Resistive Loading'," *IEEE Trans. Antennas Propag.*, vol. AP-13, no. 11 (Nov. 1965): 998.
116. E. C. Fear, J. Sill, and M. A. Stuchly, "Experimental Feasibility Study of Confocal Microwave Imaging for Breast Tumor Detection," *IEEE Trans. Microwave Theory Tech.*, vol. 51, no. 3 (March 2003): 887–892.
117. Special Issue of *IEEE Trans. Microwave Theory Tech.*, MTT-34, 1986.
118. C. H. Durney and M. F. Iskander, *Antenna Handbook*, Y. T. Lo and S. W. Lee (eds.) (New York: Springer, 1993).
119. P. K. Sneed and T. L. Phillips, "Combining Hyperthermia and Radiation: How Beneficial?," *Oncology*, vol. 5 (1991): 99–108.
120. C. C. Vernon, J. W. Hand, S. B. Field, et al, "Radiotherapy with or Without Hyperthermia in the Treatment of Superficial Localized Breast Cancer: Results from Five Randomized Controlled Trials," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 35 (1996): 731–44.
121. F. Montecchia, "Microstrip Antenna Design for Hyperthermia Treatment of Superficial Tumors," *IEEE Trans. BME*, vol. 39, no. 6 (June 1992): 580–588.
122. J. Vba, C. Franconi, F. Montecchia, and I. Vannucci, "Evanescent-Mode Applicators (EMA) for Superficial and Subcutaneous Hyperthermia," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 5 (May 1993): 397–407.
123. M. V. Prior, M. L. D. Lumori, J. W. Hand, G. Lamaitre, C. J. Schneider, and J. D. P. van Dijk, "The Use of a Current Sheet Applicator Array for Superficial Hyperthermia: Incoherent Versus Coherent Operation," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 7 (July 1995): 694–698.
124. J. Bach Andersen, A. Baun, K. Harmark, L. Heinz, P. Raskmark, and J. Overgaard, "A Hyperthermia System Using a New Type of Inductive Applicator," *IEEE Trans. Biomedical Engineering*, vol. 31(1), (1984): 21–27.
125. P. R. Stauffer, M. Leoncini, V. Manfrini, et al, "Dual Concentric Conductor Radiator for Microwave Hyperthermia with Improved Field Uniformity to Periphery of Aperture," *IEICE Trans. on Communicat.*, vol. E78-B (1995): 826–35.
126. P. F. Maccarini, H. Rolfsnes, D. Neuman, and P. Stauffer, "Optimization of a Dual Concentric Conductor Antenna for Superficial Hyperthermia Applications," *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, CA, Sept. 1–5, 2004.
127. H. Tehrani and K. Chang, "Multifrequency Operation of Microstrip Fed Slot Ring Antennas on Thin Low Dielectric Permittivity Substrates," *IEEE Trans. on Antennas and Propagation.*, vol. 50, no. 9 (Sept. 2002): 1299–1308.
128. S. Jacobsen, P. R. Stauffer, and D. G. Neuman, "Dual-Mode Antenna Design for Microwave Heating and Noninvasive Thermometry of Superficial Tissue Disease," *IEEE Trans. Biomed. Eng.*, vol. 47 (2000): 1500–9.

129. P. F. Turner, "Interstitial Equal-Phased Arrays for EM Hyperthermia," *IEEE Trans. Micr. Theory and Tech.*, vol. 34, no. 5 (May 1986): 572–578.
130. R. D. Nevels, G. D. Arndt, G. W. Raffoul, J. R. Carl, and A. Pacifico, "Microwave Catheter Design," *IEEE Trans. Biomed. Eng.*, vol. 45 (July 1998): 885–890.
131. C. Manry, S. L. Broschat, C.-K. Chou, and J. A. McDougall, "An Eccentrically Coated Asymmetric Antenna Applicator for Intracavity Hyperthermia Treatment of Cancer," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 9 (Sept. 1992): 935–942.
132. P. F. Turner, "Hyperthermia and Inhomogeneous Tissue Effects Using an Annular Phased Array," *IEEE Trans. Microwave Theory and Tech.*, vol. 32, no. 8 (Aug. 1984): 874–875.
133. P. Stauffer, J. Schlorff, R. Taschereau, T. Juang, D. Neuman, P. Maccarini, J. Pouliot, and J. Hsu, "Combination Applicator for Simultaneous Heat and Radiation," Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, Sept. 1–5, 2004.
134. Y. Kotsuka, E. Hankui, and Y. Shigematsu, "Development of Ferrite Core Applicator System for Deep-Induction Hyperthermia," *IEEE Trans. Micr. Theory and Tech.*, vol. 44, no. 10 (Oct. 1996): 1803–1810.
135. P. F. Turner, "Sigma 60–24 Prototype Test Results," BSD Medical Corporation, Internal Rep., Salt Lake City, UT, 1992.
136. J. Nadobny, H. Fähring, M. Hagmann, P. Turner, W. Wlodarczyk, J. Gellermann, P. Deuflhard, and P. Wust, "Experimental and Numerical Investigations of Feed-Point Parameters in a 3-D Hyperthermia Applicator Using Different Models of Feed Networks," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 11 (Nov. 2002): 1348–1359.
137. J. Nadobny, W. Wlodarczyk, L. Westhoff, J. Gellermann, R. Felix, and P. Wust, "A Clinical Water-Coated Antenna Applicator for MR-Controlled Deep-Body Hyperthermia: A Comparison of Calculated and Measured 3-D Temperature Data Sets," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3 (March 2005): 505–519.
138. K. S. Nikita and N. K. Uzunoglu, "Coupling Phenomena in Concentric Multi-Applicator Phased Array Hyperthermia Systems," *IEEE Trans. Microwave Theory and Tech.*, vol. 44, no. 1 (Jan. 1996): 65–74.
139. F. Bardati, A. Borrani, A. Gerardino, and G. A. Lovisolo, "SAR Optimization in a Phased Array Radiofrequency Hyperthermia System," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 12 (Dec. 1995): 1201–1207.
140. R. G. Olsen, M. B. Ballinger, T. D. David, and W. G. Lotz, "Rewarming of the Hypothermic Rhesus Monkey with Electromagnetic Radiation," *Bioelectromagnetics*, vol. 8 (1987): 183–193.
141. I. D. McRury and D. E. Haines, "Ablation for the Treatment of Arrhythmias," *Proceedings of the IEEE*, vol. 84, no. 3: 404–416.
142. A. Labonte, A. Blais, S. Legault, H. O. Ai, and L. Roy, "Monopole Antennas for Microwave Catheter Ablation," *IEEE Trans. on Microwave Theory and Techn.*, vol. 44, no. 10: 1832–1840.
143. A. Rosen et al, "Percutaneous Transluminal Microwave Balloon Angioplasty," *IEEE Trans. Microwave Theory Tech.*, vol. 38 (1990): 90–93.
144. A. J. Johansson, "Simulation and Verification of Pacemaker Antennas," Proceedings of the 25th Annual Int. Conf. of the IEEE EMBS, Cancun, Mexico, Sept. 17–21, 2003.
145. K. Guillory and R. A. Normann, "A 100-Channel System for Real Time Detection and Storage of Extracellular Spike Waveforms," *J. Neurosci. Methods*, vol. 91 (1999): 21–29.
146. T. Buchegger et al, "An Ultra-Low Power Transcutaneous Impulse Radio Link for Cochlea Implants," *Joint Ultra Wideband Systems and Technologies (UWBST) and International Workshop on UWBS 2004*, IEEE cat. no. 04EX812 (2004): 356–360.
147. K. Gosalia, G. Lazzi, and M. Humayun, "Investigation of Microwave Data Telemetry Link for a Retinal Prosthesis," *IEEE Trans. Microwave Theory and Tech.*, vol. 52, no. 8 (Aug. 2004): 1925–1932.
148. C. Furse, "Design an Antenna for Pacemaker Communication," *Microwaves & RF* (March 2000): 73–76.
149. I. J. Bahl, S. S. Stuchly, J. Lagendijk, and M. Stuchly, "Microstrip Loop Applicators for Medical Applications," *IEEE Trans. MTF* (July 1982): 1090–1093.

150. I. J. Bahl, P. Bhartia, and S. S. Stuchly, "Design of Microstrip Antennas Covered with a Dielectric Layer," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 2 (March 1982): 314–318.
151. R. D. Nevels, D. Arndt, J. Carl, G. Raffoul, and A. Pacifico, "Microwave Antenna Design for Myocardial Tissue Ablation Applications," *IEEE Antennas and Propagat. Soc. Mt. Symp.*, vol. 3 (1995): 1572.
152. C. T. Charles, "Electrical Components for a Fully Implantable Neural Recording System," Master's thesis, Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah, 2003.
153. W. G. Scanlon, N. E. Evans, G. C. Crumley, and Z. M. McCreesh, "Low-Power Radio Telemetry: the Potential for Remote Patient Monitoring," *Journal of Telemedicine and Telecare*, vol. 2, no. 4 (Dec. 1996): 185.
154. W. G. Scanlon, N. E. Evans, and Z. M. McCreesh, "RF Performance of a 418 MHz Radio Telemeter Packaged for Human Vaginal Placement," *IEEE Trans. BME*, vol. 44, no. 5 (May 1997): 427–430.
155. W. G. Scanlon, N. E. Evans, and J. B. Burns, "FDTD Analysis of Close-Coupled 418 MHz Radiating Devices for Human Biotelemetry," *Physics in Medicine and Biology*, vol. 44, no. 2 (Feb. 1999): 335–345.
156. W. G. Scanlon, J. B. Burns, and N. E. Evans, "Radiowave Propagation from a Tissue-Implanted Source at 418 MHz and 916.5 MHz," *IEEE Trans. Biomedical Engineering*, vol. 47, no. 4 (April 2000): 527–534.
157. G. C. Crumley, N. E. Evans, J. B. Burns, and T. G. Trouton, "On the Design and Assessment of a 2.45 GHz Radio: Telecommand System for Remote Patient Monitoring," *Medical Engineering and Physics*, vol. 20, no. 10 (March 1999): 750–755.
158. C. P. Yue and S. S. Wong, "On-Chip Spiral Inductors with Patterned Ground Shields for Si-based RF IC's," Center for Integrated Systems, Stanford University, 1998.
159. G. S. Brindley and W. S. Lewin, "The Sensations Produced by Electrical Stimulation of the Visual Cortex," *J. Physiol.*, vol. 196 (1968): 479–493.
160. G. E. Loeb, C. J. Samin, J. H. Schulman, and P. R. Troyk, "Injectable Microstimulator for Functional Electrical Stimulation," *Med. Biol. Eng. Comput.*, vol. 29 (1991): NS 13–NS 19.
161. B. Ziaie, M. D. Nardin, A. R. Coghlan, and K. Najafi, "A Single Channel Implantable Microstimulator for Functional Neuromuscular Stimulation," *IEEE Trans. Biomed. Eng.*, vol. 44 (Oct. 1997): 909–920.
162. Weimin Sun et al, Implantable Medical Device Microstrip Telemetry Antenna (Jan. 19, 1999): U.S. Pat. 5,861,019.
163. P. R. Troyk and M. A. K. Schwan, "Closed Loop Class E Transcutaneous Power and Data Link for Microimplants," *IEEE Trans. Biomed. Eng.*, vol. 39 (June 1992): 589–598.
164. G. M. Clark, Y. C. Tong, J. F. Patrick, P. M. Seligman, P. A. Crosby, J. A. Kuzma, and D. K. Money, "A Multi-channel Hearing Prosthesis for Profound-to-Total Hearing Loss," *J. Med. Eng., Technol.*, vol. 8 (Jan. 1984): 3–8.
165. N. de N. Donaldson and T. A. Perkins, "Analysis of Resonant Coupled Coils in Design of Radio Frequency Transcutaneous Links," *Med. Biol. Eng. Comput.*, vol. 21 (Sept. 1983): 612–626.
166. D. C. Galbraith, S. Mani, and R. L. White, "A Wide Band Efficient Inductive Transdermal Power and Data Link with Coupling Insensitive Gain," *IEEE Trans. Biomed. Eng.*, vol. BME-34 (April 1987): 265–275.
167. C. R. Pfaltz (ed.), "The University of Melbourne Nucleus Multi-Electrode Cochlear Implant," *Adv. Oto-Rhino-Laryngol.*, vol. 38 (1987): 63–81.
168. R. S. Mackay, *Bio-Medical Telemetry*, 2nd Ed. (New York: IEEE Press, 1993).
169. A. Johansson, "Wireless Communication with Medical Implants: Antennas and Propagation," PhD dissertation, Lunds Universitet, 2004.
170. M. D. Amundson, J. A. Von Arx, W. J. Linder, P. Rawat, and W. R. Mass, Circumferential Antenna for an Implantable Medical Device (2002): U.S. Pat. 6,456,256
171. L. Griffiths, "Analysis of Wire Antennas for Implantation in the Body," Master's thesis, Utah State University, Logan, UT, 2002.

172. M. Sun, M. Mickle, W. Liang, Q. Liu, and R. J. Scلابassi, "Data Communication Between Brain Implants and Computer," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2 (2003): 189–192.
173. P. Soontornpipit, C. M. Furse, and Y. C. Chung, "Design of Implantable Microstrip Antennas for Communication with Medical Implants," *IEEE Trans. MTT*, vol. 52, issue 8 (Aug. 2004): 1944–1951.
174. J. Kim and Y. Rahmat-Samii, "Implanted Antennas Inside a Human Body: Simulations, Designs, and Characterizations," *IEEE Trans. MTT*, vol. 52, issue 8 (Aug. 2004): 1934–1943.
175. K. Gosalia, J. Weiland, M. Humayun, and G. Lazzi, "Thermal Elevation in the Human Eye and Head Due to the Operation of a Retinal Prosthesis," *IEEE Trans. Biomedical Engineering*, vol. 51, no. 8 (Aug. 2004).
176. M. Ghovanloo and G. Lazzi, "Transcutaneous Magnetic Coupling of Power and Data," *Wiley Encyclopedia of Biomedical Engineering*, M. Akay (ed.) (New York: John Wiley & Sons, 2006).
177. C. Furse, G. Lazzi, and O. P. Gandhi, "Dipoles, Monopoles, and Loop Antennas," (invited chapter), *Modern Antennas*, C. Balanis (ed.) (Springer, est. publish date 2007).
178. P. Soontornpipit, C. M. Furse, and Y. C. Chung, "Miniaturized Biocompatible Microstrip Antenna Using Genetic Algorithm," *IEEE Trans. Antennas and Propagat.* (June 2005): 1939–1945.
179. C. A. L. Bassett, "Bioelectromagnetics in the Service of Medicine," *Electromagnetic Fields and Interactions*, Martin Blank, (ed.), American Chemical Society (1995): 262–275.
180. J. C. Weaver, R. Langer, and R. O. Potts, "Tissue Electroporation for Localized Drug Delivery," *Electromagnetic Fields and Interactions*, Martin Blank (ed.), American Chemical Society (1995): 301–316.
181. F. Sterzer, "Microwave Medical Devices," *IEEE Microwave Magazine* (March 2002): 65–70.
182. E. C. Burdette, F. L. Cain, and J. Seals, "In Vivo Probe Measurement Technique for Determining Dielectric Properties at VHF Through Microwave Frequencies," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (April 1980): 414–427.
183. T. K. Bose, A. M. Bottreau, and R. Chahine, "Development of a Dipole Probe for the Study of Dielectric Properties of Biological Substances in Radio Frequency and Microwave Region with Time-Domain Reflectometry," *IEEE Trans. on Instr. and Meas.*, vol. IM-35, no. 1 (March 1986): 56–60.
184. L. Aamodt, M. Manwaring, and K. Manwaring, "In Vivo Brain Tissue Water Measurement," *Proceedings 12th IEEE Symposium on Computer-Based Medical Systems* (June 18–20, 1999): 130–135.
185. H. Bassen and G. Smith, "Electric Field Probes—A Review," *IEEE Transact. on Antennas and Propagat.*, vol. AP-31, no. 5 (Sept. 1983): 710–718.
186. Bassen, "Electric Field Probes for Cell Phone Dosimetry," *Proc. 19th International Congress IEEE/EMBS Society*, Chicago, Illinois, Oct. 30–Nov. 2, 1997.

Chapter 39

Automobile
Antennas

Louis L. Nagy
Delphi Research Labs

CONTENTS

39.1	INTRODUCTION	39-2
39.2	AM ANTENNAS	39-2
39.3	AM/FM MAST ANTENNAS	39-7
39.4	AM/FM WINDOW GLASS ANTENNAS	39-10
39.5	AM/FM PANEL ANTENNAS	39-14
39.6	FM DIVERSITY ANTENNAS	39-15
39.7	RKE AND TPMS ANTENNAS	39-16
39.8	UHF/VHFTV ANTENNAS	39-17
39.9	MOBILE TELEPHONE ANTENNAS	39-18
39.10	GPS ANTENNAS	39-19
39.11	SDARS ANTENNAS	39-26
39.12	EMERGING TECHNOLOGIES	39-30
39.13	ANTENNA MEASUREMENTS	39-31

39.1 INTRODUCTION

Automobile communications are rapidly expanding with the emergence of many new telecommunication technologies (e.g., XM Radio and Sirius Satellite Radio, WiFi, Bluetooth, mobile TV, Dedicated Short Range Communication, etc.), as illustrated in Figure 39-1. This chapter reviews the antenna technologies that are being used for these various communication systems. In addition, this chapter reviews the antenna requirements as established by vehicle manufacturers in terms of radiation/reception characteristics, cost, aesthetics, reliability, functionality, repairability, design flexibility, RFI/EMC immunity, and buildability (i.e., the ability to place the antenna into the production process of an automobile assembly plant).

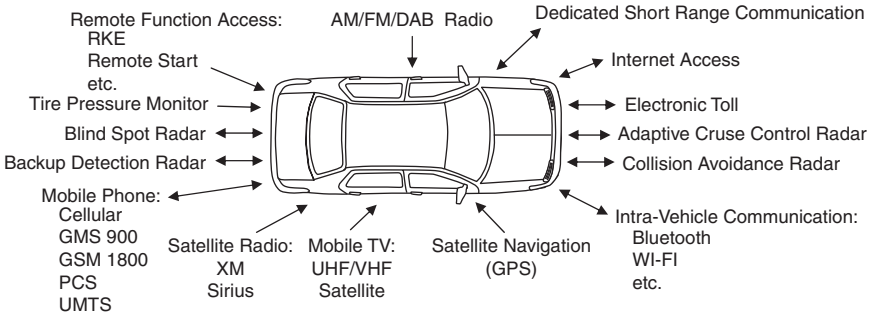


FIGURE 39-1 Mobile communication services

39.2 AM ANTENNAS

The AM broadcast frequency band is from 0.530 to 1.710 MHz with its wavelengths varying from 566 to 175 m. These wavelengths are considerably longer than the maximum dimension of any automobile antenna system (i.e., antenna element and vehicle body structure). Thus, the AM car antenna is considered to be a very small antenna.¹ Illustrated in Figure 39-2a is the geometry of an automobile at a height h (where $h \ll \lambda$) above the earth's surface, and its simple electromagnetic model is shown in Figure 39-2b. The AM impedance path through each wheel-bearing and tire structure (relative to the earth's surface) is modeled by a lump

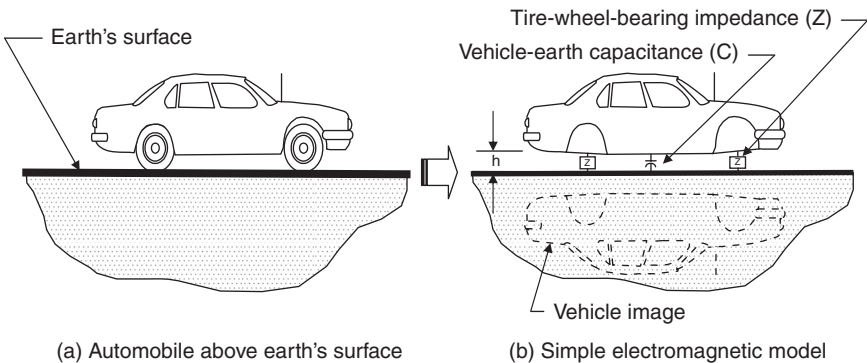


FIGURE 39-2 AM frequency model for vehicle above earth's surface

impedance element (Z) while the vehicle-earth capacitance is modeled by a lump capacitor element (C). This capacitor can be approximated by using image theory and representing the undercarriage of the vehicle as a flat planar conductor at a height h above the earth's surface. The dielectric characteristics for various soil conditions (ranging from dry to 16.8 percent moisture) were measured at 1 MHz and found to range from a relative dielectric constant (ϵ_r) of 2.5 to 20 with a loss tangent ($\tan\delta$) of 0.017 to 4.0.² Dry soil has a low loss tangent and can be treated as a lossy dielectric, whereas moist soil has a high loss tangent and can be treated as a lossy conductor. The vehicle-earth capacitance is known to have a significant effect on the AM reception of the vehicle. This capacitance, for a medium-sized automobile, was measured to be approximately 250 pF.³

Early Automobile Antennas

Factory-installed vehicle radio systems began to appear in 1923 with the crude addition of a rather large receiver, antenna, battery, and speaker.⁴ The actual integration of the radio system into the vehicle had to wait until the early 1930s when vehicle manufacturers provided an in-dash radio with a factory-installed antenna. At that time, many of the exterior vehicle panels were made of composite materials (wood, canvas, glass, etc.) that allowed the development of a variety of hidden antenna systems: the Under-Car, Spare Wheel, Running Board, Rooftop Screen (commonly known as the Chicken Wire Antenna), Rooftop Wire (a long wire threaded into the fabric of the roof), and Retractable Mast Antenna (see Figure 39-3). Of these early systems, the rooftop screen antenna gave best reception due to its large size, height above ground, and location away from the car engine.

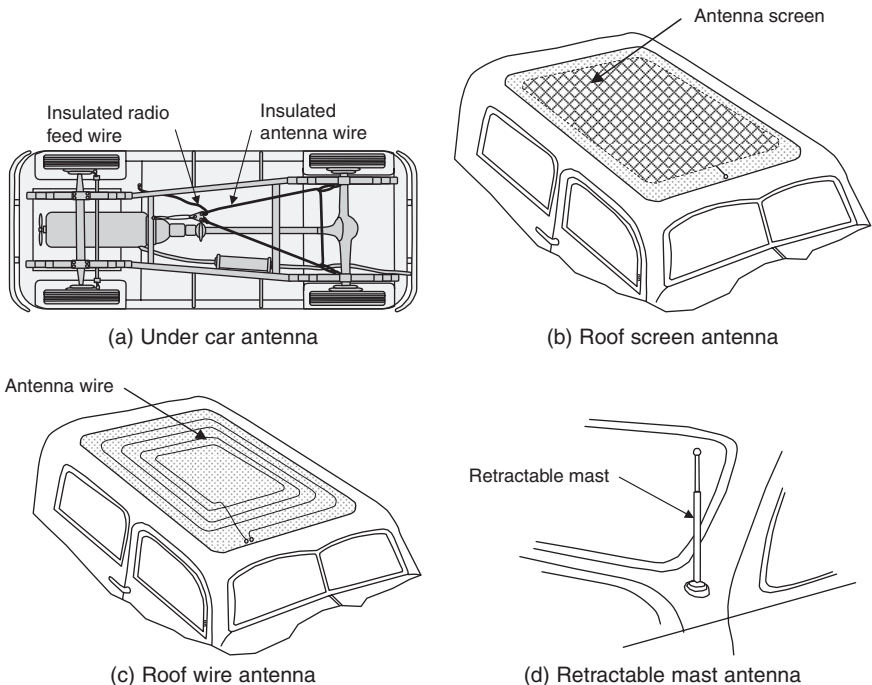


FIGURE 39-3 Early AM vehicle antenna systems

However, a major setback for this antenna came in 1934 in the form of the all metal “Turret Top” roof manufactured by General Motors, which was quickly adopted by other automobile manufacturers.⁵ The advent of the all-metal roof resulted in the mast antenna becoming the dominant vehicle antenna until the end of the 20th century. The length of these early mast antennas generally ranged from three to five feet. The longer the antenna element, the higher the AM gain, but the more unsightly it became.

Mast Antennas

A very short monopole antenna mounted on a perfectly conducting infinite ground plane has a radiation resistance given by the following expression (i.e., when the wavelength is much longer than the antenna length, antenna length is much longer than its radius, and current distribution is assumed to be triangular in shape⁶):

$$R_A = 40 (\pi \ell / \lambda)^2 \quad (39-1)$$

where ℓ is the length of the antenna, λ is the wavelength, and $\ell < \lambda/30$. This formula can be used to obtain a good approximation for R_A for antenna lengths up to $\lambda/8$. The reactance for a very short monopole is capacitive with values given by the following expressions⁷:

$$X_A \approx -60[\ln(\ell/a) - 1] / [2\pi\ell/\lambda] (\Omega) \quad (39-2)$$

and

$$C_A \approx [2\pi\ell/\lambda] / (60\omega[\ln(\ell/a) - 1]) (\text{F}) \quad (39-3)$$

where X_A is the antenna reactance, C_A is the antenna capacitance, a is the radius of the antenna element, ω is the angular frequency ($2\pi f$), f is the RF operating frequency, and $a \ll \ell$.

A typical automobile mast antenna has an average height (ℓ) of about 30 inches (0.76 m), and an average diameter ($2a$) of about 1/8 inch (3.2 mm). When this mast is mounted on a perfectly conducting infinite ground plane and operated at 1 MHz, it will have an ℓ/λ of 0.0025, a/ℓ of 0.0021, radiation resistance of 0.0025 Ω , reactive impedance of $-19.7 \text{ K}\Omega$, and capacitance of 8.1 pF. It should be noted that the reactive impedance is several orders of magnitude greater than the radiation resistance. Thus, this antenna can be viewed as a capacitor probe.

The radiation efficiency for a small monopole antenna is given by the following equation:

$$e = R_A / (R_A + R_{\text{Ohmic}}) \quad (39-4)$$

where e is the antenna radiation efficiency, and R_{Ohmic} is the ohmic loss of the antenna element. The typical automobile mast element is constructed of materials having high conductivity (e.g., Ni-Cr stainless steel with a conductivity $\approx 10^6/\Omega \text{ m}$) that results in the antenna currents being confined to the thin outer layer of the conductor. For this current condition, the following equation can be used to determine the ohmic loss of the antenna element⁸:

$$R_{\text{Ohmic}} = \ell R_s / 3P \quad (39-5)$$

where $R_s = (\omega\mu/2\sigma)^{1/2}$ is the high-frequency surface resistance, P is the circumference ($2\pi a$), μ is the permeability, and σ is the conductivity of the antenna element. A Ni-Cr stainless steel mast antenna with a height of 30 inches, diameter of 1/8 inch, operated at 1 MHz,

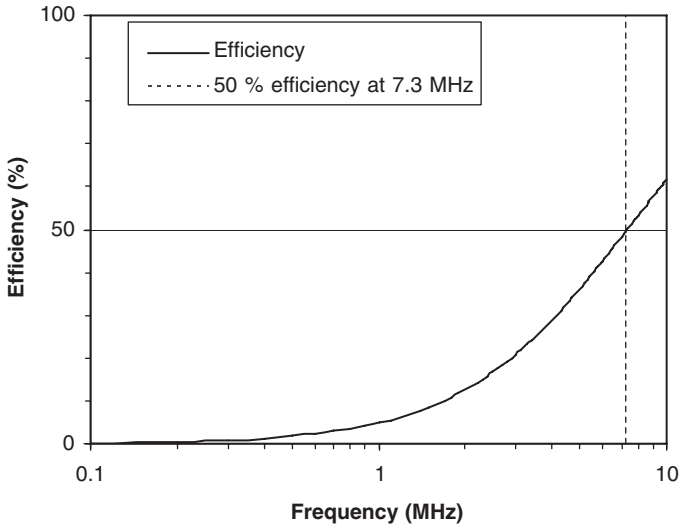


FIGURE 39-4 Efficiency of a very small monopole antenna

and mounted on a perfectly conducting infinite ground plane has an R_A of $0.0025\ \Omega$, R_s of $0.002\ \Omega$, R_{Ohmic} of $0.050\ \Omega$, and radiation efficiency of 4.7 percent. A plot of the efficiency versus frequency is contained in Figure 39-4 for the frequency range from 0.1 to 10 MHz. It should be noted that this simple antenna model does not take into account any of the effects associated with the vehicle body structure. The selection of the coaxial cable, from the antenna to the radio, is important in achieving quality AM reception. Shown in Figure 39-5 is a simple electrical model for the automobile's AM antenna-radio system (i.e., a Norton equivalent circuit with the negligible resistance and inductance components removed), where I_{AN} = induced short-circuit antenna current, I_R = current passing through the radio capacitor, C_A = antenna capacitance (including vehicle structure effects), C_C = coaxial cable capacitance, and C_R = radio capacitance.

The three modeled capacitors form a current divider circuit with the radio current given by $I_R = I_{\text{AN}} (C_R / (C_R + C_C + C_A))$. The input capacitance of the radio is about 75 pF, capacitance of the 30-inch mast antenna is about 10–15 pF, and antenna base-mount capacitance is about 15 pF. The length of coaxial cable for a front-mounted fender mast antenna is approximately 5 feet, while the length for a rear-mounted fender mast can be over 20 feet.

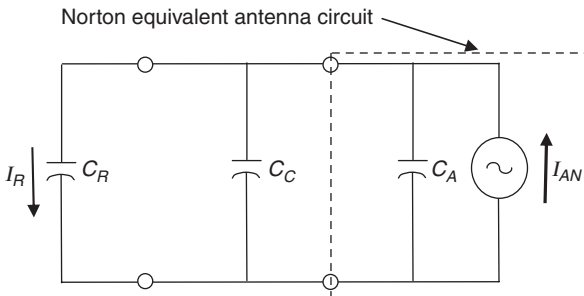


FIGURE 39-5 Norton equivalent circuit for vehicle AM antenna-radio system

Because of the increase in cable capacitance, the front-mounted mast antenna system (including its associated coaxial cable) has an effective AM gain several dB greater than a rear-mounted mast antenna system (assuming the same type coaxial cable is used for both antennas). In order to avoid excessive AM signal loss, the rear-mounted antenna requires the use of a very low-capacitance/length cable (<10 pF/ft), or possibly a pre-amplifier. A mast mounted in the center of the roof has the best reception because of its increased overall height and acceptable cable length (about ten feet). Unfortunately, a 30-inch roof mast antenna is viewed as being incompatible with vehicle aesthetics and more susceptible to damage (by low tree branches, garage door height clearance, car wash facility, etc.). Table 39-1 contains a chart that compares common automobile antennas. The fender-mounted mast was the dominant vehicle antenna until the end of the 20th century.

A retractable mast antenna consists of several segmented elements (making electrical contact with each other by means of sliding contacts), a storage cylinder for the retracted elements (located beneath the sheet metal of the fender), and an antenna feed system attached to the cylinder. It is viewed as a partially hidden antenna in that it can be lowered when the radio is not in use. Its AM characteristics are much like those of the fixed mast antenna, but it has a slightly lower AM gain resulting from the additional shunning capacitance of the cylinder-feed system.

TABLE 39-1 AM/FM Antenna Comparison Chart

Attribute	Antenna Type								Active Rear Window	Active Diversity Window
	Fixed Mast	Power Mast	Wire Windshield	Film Windshield	Roof Slot	Decklid Slot	Rear Window	Fair/ Poor		
AM Gain	Good	Good	Fair	Good	Good	Good	Fair/ Poor	Good	Good	
FM Gain	Good	Good	Fair	Good	Good	Fair	Fair	Good	Good	
FM Omni- Azimuth Patterns	Good	Good	Fair/poor	Fair	Fair	Fair/ Poor	Fair/ Poor	Fair	Good	
Electrical Noise Immunity	Good	Good	Fair/Poor	Good/Fair	Good	Fair	Fair	Fair	Fair	
Signal Overload Resistance	Good	Good	Good	Good	Good	Good	Good	Fair	Fair	
Aesthetics	Poor	Fair	Good	Good	Good	Good	Good	Good	Good	
Reliability	Fair	Poor	Good	Good	Good	Good	Good	Good	Good	
Manufacturability	Good	Fair	Good	Fair	Good	Fair	Good	Fair	Fair	
Serviceability	Fair	Poor	Good	Good	Good	Fair/ Poor	Good	Good	Good	Fair
Weather Resistance	Fair	Poor	Good	Good	Good	Good	Good	Good	Good	
Weight	Fair	Poor	Good	Good	Good	Good	Good	Good	Good	
Wind Noise Generation	Fair/ Poor	Fair/ Poor	Good	Good	Good	Good	Good	Good	Good	
Warranty	Fair	Poor	Good	Good	Good	Fair	Fair	Fair	Fair	
Vandalism Resistance	Poor	Fair	Good	Good	Good	Good	Good	Good	Good	
Damage Resistance ¹	Poor	Fair/ Poor	Good	Good	Good	Good	Good	Good	Good	
Cost	Good	Poor	Fair	Fair*	Fair	Fair	Good	Fair	Fair/Poor	

¹Car wash facilities, tree branches, etc.
 *Requires conductive solar windshield film

Trunk Lid Antennas

The 1953 to 1955 Corvette had a unique hidden trunk lid antenna. This antenna took the form of a galvanized-steel screen⁹ embedded into the bottom surface of its fiberglass trunk lid. The center conductor of the radio coaxial cable was soldered to the screen (before it became a permanent part of the trunk lid). The outer shield of the cable was grounded midway between the antenna and radio with a ground strap that was attached to the underside of the body frame. This antenna had acceptable AM reception, but was replaced in 1956 by a retractable rear-mounted fender mast that cost less and was easier to service.

39.3 AM/FM MAST ANTENNAS

In the process of establishing the FM broadcast service (i.e., from 87.7 to 107.9 MHz), the Federal Communications Commission (FCC) determined that urban (man-made) RF electrical noise was predominantly vertically polarized (by about 10 to 20 dB). In an attempt to achieve the best quality of reception (i.e., signal/noise), the FCC initially selected horizontal polarization for FM broadcasting. However, the dominant vehicle antenna during the 1960s and until the end of the 20th century was the vertical mast antenna. Thus, the automobile mast antenna was being operated in a cross-polarized mode that lowered its reception gain by about 10 dB (i.e., the average horizontal polarization gain of a vertical mast is about 10 dB below its average vertical polarization gain). This problem was eventually addressed by the FCC by allowing most FM broadcast stations to transmit both polarizations. The result is that some broadcast stations transmit signals using both polarizations at equal power levels, some using circularly polarized signals, some using vertically polarized signals at lower power levels, some using only horizontally polarized signals, and a lesser number using only vertically polarized signals.¹⁰ Thus, the FM reception environment contains a mixture of both vertically and horizontally polarized signals having different power levels depending on the particular broadcast station and signal propagation conditions. To help determine real-world polarization characteristics of the FM broadcast environment, a limited polarization measurement survey was conducted using randomly selected FM broadcast signals (52) from a variety of suburban and rural locations. The result of this survey is plotted in Figure 39-6.

This data indicates that the majority of the measured signals are elliptically polarized with horizontal elliptically polarized signals (i.e., their horizontal component being greater than their vertical component) slightly more numerous (35) than vertical elliptically polarized signals (i.e., their vertical component being greater than their horizontal component) (27). Four signals were judged to be circularly polarized (i.e., differences between the horizontal and vertical polarized components being less than ± 1 dB while their phase differences being within $90^\circ \pm 10^\circ$). Six signals were judged to be horizontally polarized (i.e., differences between the horizontal and vertical polarization components being greater than 15 dB).

FM vehicle antennas should receive broadcast signals equally well from any direction. Hence, these antennas should be designed to have omnidirectional patterns in their azimuth plane. However, it has been demonstrated that the FM characteristics of an automobile antenna are very dependent on the structure of the vehicle. This is due to the various dimensions of an automobile (including its window openings) being comparable to the wavelengths of FM signals. Because of this vehicle body dependence, automobile manufacturers state their antenna specifications (i.e., gain, VSWR, max/min pattern values, etc.) for an antenna mounted on the intended production vehicle. These antenna specifications can vary from one automobile manufacturer to another.

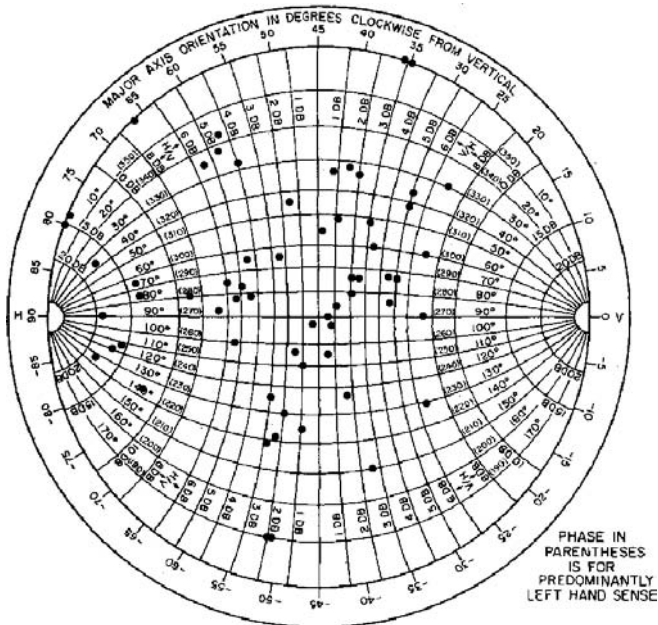


FIGURE 39-6 Measured polarization characteristics of FM broadcast signals

The complex interaction of the antenna with the vehicle body prevents analytical techniques from being used in the design process of these antennas. The present limitations of EM simulation software and computer capabilities restrict their use to only approximate solutions. Thus, an iterative FM antenna design process is used that consists of the following steps:

1. Design a prototype antenna using simulation techniques and/or hands-on experience.
2. Build a reasonable facsimile of this antenna.
3. Mount the antenna on the intended production vehicle (or a reasonable mock-up of it).
4. Measure its RF characteristics.
5. Determine if the antenna meets the design goals and, if not, repeat the process.

This type of design approach is very time-consuming, costly, and demanding on supplier resources. However, there is an expectation that future advancements in EM simulation software and more powerful computers will provide the tools needed to more accurately design vehicle antenna systems and greatly reduce the need for an iterative design process.

Passive Mast Antenna

The AM/FM mast system is basically a monopole antenna having a height of about $\lambda/4$ at 100 MHz (i.e., about 30 inches). Placement of the mast antenna in the middle of the roof gives best gain for vertically polarized signals with nearly omnidirectional reception patterns. These patterns become more directional as the antenna is moved near the front, rear, and/or side edges of the roof. A tilted roof mast antenna has an increased

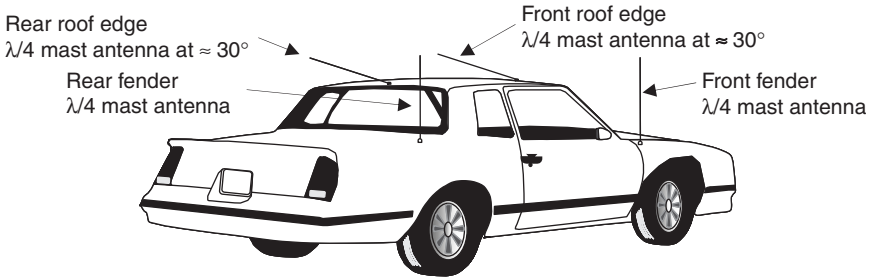


FIGURE 39-7 Common fixed AM/FM mast antenna locations

gain for horizontally polarized signals and reduced gain for vertically polarized signals (see Figure 39-7). Placement of the mast on the front or rear fender will reduce AM/FM gain and increase FM pattern directionality. However, the aesthetic benefit of a fender-mounted antenna generally outweighs the minor degradation in its reception performance. In addition, moving a fender mast closer to or further away from the vehicle's pillars will affect its FM characteristics.

The retractable AM/FM mast antenna consists of multiple sliding elements, or a single sliding element (see Figure 39-8). The multiple-element antenna is generally mounted on the passenger side of the front or rear fender, while the single-element antenna is generally mounted at the driver side of the roof pillar where it can be manually raised and lowered by the driver. The AM/FM characteristics of these antennas are similar to those of the fixed mast, with the AM gain being slightly lower because of the additional capacitance from the cylindrical-tube feed system.

The power mast antenna consists of several sliding antenna elements, an electric motor for raising and lowering these elements, a cylindrical tube for storing the retracted elements (located beneath the sheet metal of the vehicle), and an antenna feed system, generally attached to the side of the storage cylinder. Common locations for this type of antenna are on the passenger side of the front or rear fender. The power mast is raised when the vehicle radio is turned on and lowered when the radio is turned off. Hence, these antennas

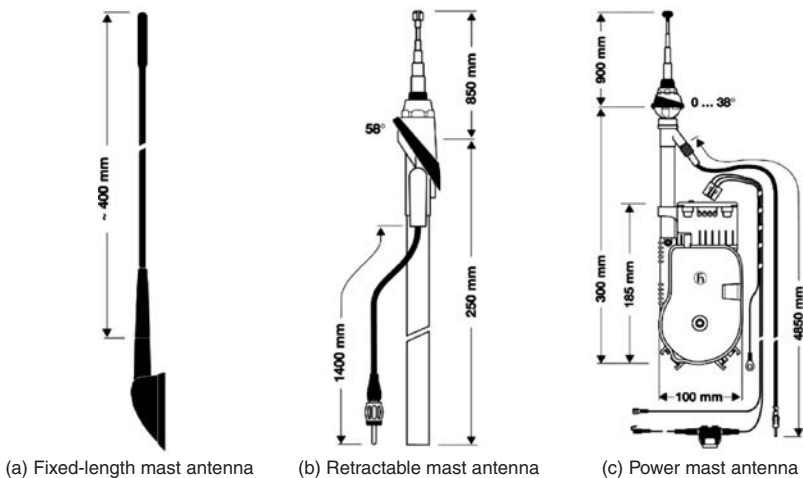


FIGURE 39-8 Production mast antennas (Courtesy of Hirschmann Car Communication, Inc.)

are hidden when not in use. The AM/FM characteristics of these antennas are similar to those of fixed mast with a slightly lower AM gain resulting from the additional shunning capacitance of the storage cylinder and antenna motor drive mechanism.

Active Mast Antennas

The active mast antenna became commercially available in the early 1970s. The fundamental advantage of this type of antenna is a reduced physical length (e.g., 16 inches), which is achieved by adding an AM/FM pre-amplifier module at its feed port. This module consists of two separate pre-amplifiers: one for the AM band and another for the FM band. Pre-amplifiers are susceptible to strong signal degradation and must be designed to be highly linear devices with signal-level conditioning circuitry to reduce overload, cross-modulation, and intermodulation effects.

The AM pre-amplifier is basically a buffer amplifier with a voltage gain of unity, high input impedance (Z_{in}), low output impedance (Z_{out}), and power gain of Z_{in}/Z_{out} . To reduce antenna loading, the input impedance of the amplifier is much greater than the AM reactive impedance of the antenna. To reduce coaxial cable loss, the output impedance of the amplifier is much smaller than the reactive impedance of the coaxial cable. The FM pre-amplifier generally has a gain of several dB, and a good impedance match with the coaxial cable.

39.4 AM/FM WINDOW GLASS ANTENNAS

The dimensions of the vehicle and its enclosed compartments (i.e., passenger, trunk, and engine compartment) are electrically small compared to the wavelengths of AM broadcast signals. Therefore, an AM signal that propagates into these compartments is significantly attenuated, and the reception characteristics of an AM antenna placed within these compartments will be greatly reduced. However, the placement of an antenna in a nonconductive outer panel of the vehicle can significantly circumvent this attenuation phenomenon, as was demonstrated by the early AM roof screen, roof wire, and trunk lid antennas. This signal attenuation phenomenon is less of a problem at FM frequencies where the dimensions of the vehicle and its enclosed compartments are comparable to the wavelengths of the FM signals. At the beginning of the 21st century, the placement of AM/FM antennas in vehicle windows became a very common practice. This mounting technique is analogous to a cavity-back antenna system where the vehicle cavity is below the cutoff frequency for AM, but potentially near resonance at FM frequencies. The characteristics of these window antennas are very dependent on the vehicle body structure, and this strong dependency requires that these antennas be designed specifically for the vehicle on which they are to be used. There are three fundamental types of AM/FM window antennas:

- Fine-wire in-glass antenna
- Frit on-glass antenna
- Transparent conductive film in-glass antenna

Fine-Wire In-Glass Antennas

The windshield is formed by a heat laminating process that binds together two pre-cut glass sheets (i.e., an outer and inner sheet) with a sheet of polyurethane adhesive material (sandwiched between the two glass sheets). The fine-wire windshield antenna

required an additional manufacturing step of adhering fine wires to the polyurethane sheet before lamination. These wires had a very small diameter (in order to successfully insert them in between the glass sheets) and were coated with a dark dull color (to make them less visible).

The GM passive wire windshield antenna was the first production wire in-glass antenna, which was introduced in the 1969 Pontiac Grand Prix.¹¹ It was constructed using two 0.003-inch-diameter wires that were formed into inverted L-shaped geometries, as shown in Figure 39-9. The inclined vertical wire elements were designed to capture AM broadcast signals, while the horizontal wire elements were designed to act as load elements (i.e., capacitively coupled with the front edge of the roof). Its AM capacitance was approximately 40 pF and the AM gain was several dB below the 30-inch mast antenna. For FM operation, the wire elements were similar to bent $\frac{1}{2}\lambda$ mast antennas, with resonance achieved by selecting appropriate lengths for the inclined vertical and horizontal elements, gap width between the inclined vertical elements, and gap widths between the horizontal elements and the front edge of the roof. The FM impedance was approximately 210 Ω and the FM gain was several dB below the 30-inch ($\lambda/4$) mast antenna. In addition, the FM antenna patterns of the windshield antenna were more directional than those of the mast antenna.

The antenna wire was connected to the radio's coaxial cable at a small half-oval shaped opening that was designed into the outer glass sheet of the windshield at its bottom center edge. The two antenna wires and a small pigtail lead were fastened to a thin copper strip that was attached to the polyurethane sheet material. The small half-oval opening was hermetically sealed to protect the connection from weathering effects. The center conductor of the coaxial cable was connected to the pigtail lead, while its outer shield was attached to a DC grounding strap. This feed location placed the antenna terminals very close to the engine and instrument panel (IP), exposing it to high levels of electrical noise. The result was that the quality of its AM reception (signal-to-noise ratio, or SNR) was significantly lower than that of a front-mounted fender mast. To help improve reception, extra RF noise-suppressing techniques were used for the engine, a grounded sheet of aluminum foil was positioned beneath the IP padding, and grounding straps were mounted between the hood and engine firewall.

Other types of passive fine-wire antenna systems were investigated but did not achieve the same acceptance as this windshield antenna. However, AM and FM pre-amplifiers have been used to improve the gain characteristics of these antennas. This approach was employed for a number of European vehicles starting in the mid-1980s.¹² These antennas were constructed in a similar fashion as that of the passive windshield antenna. An example of this type of active windshield antenna is shown in Figure 39-10. The embedded wire had an inverted U-shape geometry with wire lengths and gap spacings selected for best reception. The AM and FM pre-amplifiers were located at the base of the pillar.

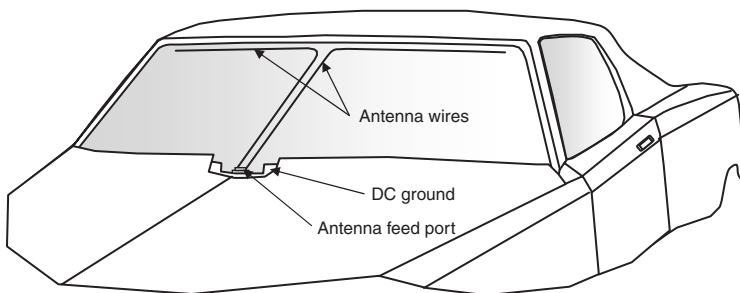


FIGURE 39-9 AM/FM wire windshield antenna

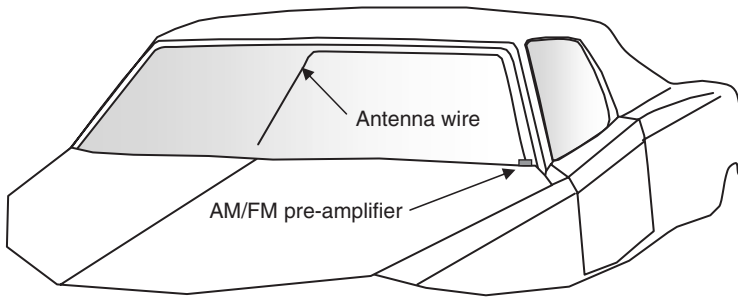


FIGURE 39-10 AM/FM active windshield antenna

Frit On-Glass Antennas

The first frit on-glass antenna was located in the rear window of the automobile, where the rear window defogger lines were used as the antenna elements with suitable RF interface circuitry placed between the defogger grid and the coaxial cable. The geometries of these antennas were dictated by the design requirements for the rear window defogger. The defogger grid consists of equally spaced parallel lines connected at each end to frit bus bars. The defogger frit material is a paste-type mixture of silver, ceramic, and a vitrifying binder material that is "silk-screened" on the inside surface of the rear window and heat treated to fuse the frit paste to the glass surface. The result is a low-resistance material having an extremely strong bond to the glass. The resistance of the defogger line is controlled by its width, which is usually about 0.8 mm with a resistance per unit length of approximately $0.17 \Omega/\text{inch}$. The bus bars are significantly wider (about 12.7 mm) to provide a low-resistance feed path for the DC heater current. Experience has demonstrated that these finite conducting frit lines can be used for AM/FM antenna purposes.¹³ However, the early passive rear defogger antennas (e.g., see Figure 39-11) were unacceptable because of their poor AM/FM gain and impedance mismatch.

A different rear window antenna geometry that uses additional frit lines located in the region above the defogger grid (as shown in Figure 39-12) was determined to have acceptable AM/FM performance.¹⁴ Its AM gain was several dB below the gain of a mast antenna, but its SNR was found to be equal to both the mast and active backlite antennas (when mounted on this same type vehicle). It was determined that vehicle radiated noise was the dominant noise received by these antennas in fringe reception areas. This was validated by reception

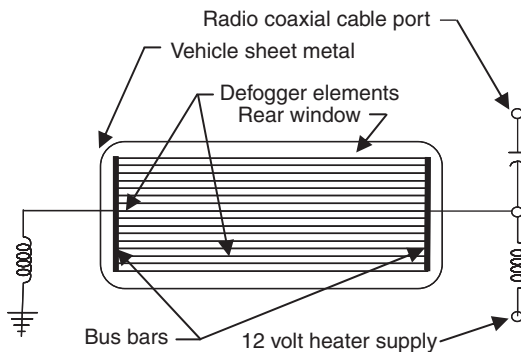


FIGURE 39-11 Early passive rear window antenna

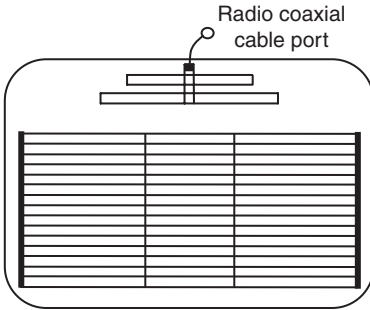


FIGURE 39-12 New passive rear window antenna

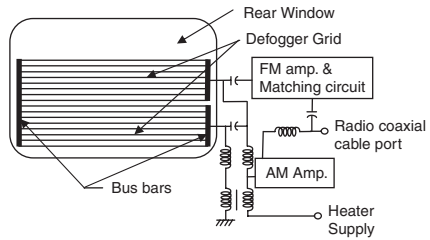


FIGURE 39-13 Early active rear window antenna

tests conducted in a large number of AM fringe broadcast environments that were located throughout the continental United States. The only noticeable difference in AM reception was the radio seek function, which was corrected by changing the seek reference level of the radio from 62 μ volts to 18 μ volts.¹⁵ This antenna was placed into production in 2001.

Advancements in AM/FM pre-amplifiers and impedance matching circuits have significantly improved the performance of the rear window antennas. This has resulted in rear window antennas being the dominant AM/FM type of automobile antenna at the beginning of the 21st century. Shown in Figure 39-13 is one of the earlier active defogger antennas.¹⁶ Like the mast antenna, the defogger antenna is a small capacitive type AM antenna with its capacitance ranging from 75 to 400 pF depending on its geometry and proximity to the vehicle sheet metal. Its AM impedance is high (e.g., a reactive impedance of 2 k Ω for an antenna with a 75 pF capacitance at 1 MHz) compared to the low impedance of the defogger heater supply circuit. To AM isolate this antenna from the defogger heater supply circuit, a bifilar choke was placed between the antenna tap-off point of the grid and heater supply connection. An AM buffer amplifier was placed at the antenna port with its output connected to the radio's coaxial cable via an RF choke. An FM choke was placed in-series with the bifilar choke for FM isolation purposes. Acceptable FM reception is achieved with the use of an impedance matching circuit and a low noise pre-amplifier having a gain of several dB. This antenna was popular in Europe where fringe reception was not as critical an issue as it is in the United States. The FM reception was reported to be good for horizontal polarization (for application in Europe) and vertical polarizations (for application in the United Kingdom).

The next evolutionary step for the active rear window antenna was to add some additional frit patterns above and/or below the defogger grid. The addition of these separate frit lines eliminated the need for the bifilar choke, thereby eliminating some rather heavy components.¹⁷ An AM pre-amplifier, FM pre-amplifier, and matching circuit module was added, as shown in Figure 39-14. Many rear window antenna geometries have been placed into production by automobile manufacturers with no clear indication as to which provides best reception. The search for a universal antenna geometry that works well on all vehicle body styles does not yet exist, and may have to wait until adequate simulation software becomes available. The frit side-window antenna is another form of active antenna that is a very common on-glass antenna for SUVs.

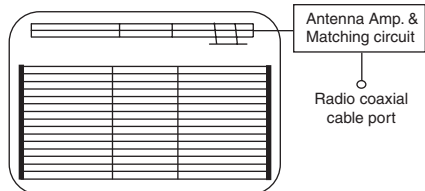


FIGURE 39-14 Active rear window antenna

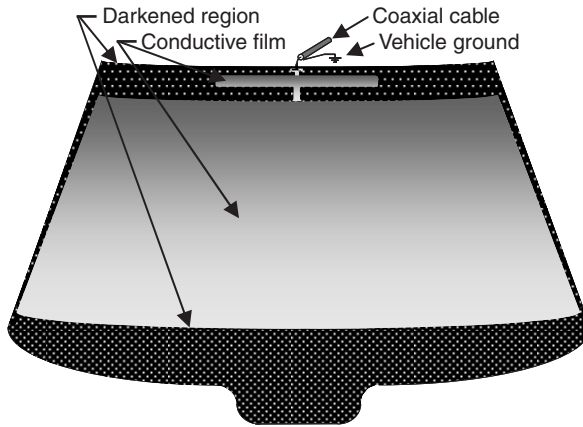


FIGURE 39-15 Solar-ray windshield antenna

Transparent Film In-Glass Antennas

The transparent conductive film antenna consists of a shaped film (with a sheet resistance of about 4 ohms per square) placed between the windshield's inner and outer glass sheets.¹⁸ The film consists of multiple layers of an antireflective metal oxide layer (e.g., oxide of zinc and tin), an infrared reflective layer (e.g., silver), and a primer layer (e.g., titanium).¹⁹ These multiple layers are sputtered on the inside surface of the outer glass sheet and are used for both solar control (i.e., infrared heat-load reduction) and antenna purposes. The film geometry has an upper horizontally elongated principle element parallel to the upper edge of the windshield opening with a length of a quarter wavelength at 100 MHz (see Figure 39-15). A small vertical film element connects the lower edge of this principle element to the upper edge of a much larger lower secondary element that covers the entire nondarkened region of the windshield. The upper horizontal edge of the lower element is located at a distance below the lower edge of the upper principle element so that coupling between them is minimal. The side edges and lower edge of the lower element form a parasitic slot transmission line with the sheet metal of the vehicle. The effective length of this slot transmission line is a multiple of half wavelengths (i.e., at 100 MHz). This lower secondary element is used as an impedance matching device for the principle element at FM frequencies and to increase AM gain. The antenna feed port is located at the upper center edge of the windshield, and can be either directly connected or capacitively coupled (by means of an added frit grid on the inside surface of the windshield) to the principle element.²⁰ The AM gain for this antenna is greater than or equal to that of the mast antenna (depending on the antenna feed coupling technique) while its FM gain and pattern characteristics are comparable to those of the mast antenna.

39.5 AM/FM PANEL ANTENNAS

The use of nonconductive vehicle body panels (i.e., other than the glass windows) for antenna purposes was investigated by a number of groups over the past decades. These panels included the roof, trunk lid, hood, side body section, bumper, roof rack, spoiler, etc. In general, these types of antennas have poor antenna characteristics and are difficult to

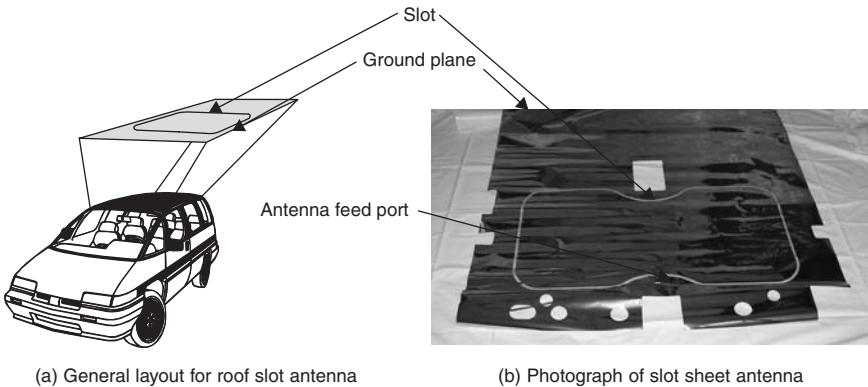


FIGURE 39-16 Roof slot antenna

incorporate into the vehicle production process. In addition, replacing a sheet metal panel with a composite panel has proven difficult for vehicle manufacturers because of fit-and-trim difficulties, “orange-peel” paint-finish concerns, and cost.

One successful application of a panel antenna was the GM APV roof slot antenna that was placed into production in 1992.^{21,22} This vehicle had composite outer panels that permitted the placement of an annual slot sheet antenna in between its roof panel and headliner (see Figure 39-16). The antenna consisted of a copper film (one micron thick) sputtered on a 1-mil polyester sheet that included an etched 6-mm-wide slot gap having a slot perimeter of about 3.0 m. A thin protective overcoat was placed on top of the copper film to mitigate atmospheric corrosion effects. The antenna feed port was located at the front center point of the slot, with the inner conductor of the coaxial cable connected to the inner copper portion of the antenna, and its outer shield connected to the outer copper portion of the antenna. The AM gain for this antenna was several dB greater than a front fender mast antenna, while its FM characteristics were comparable to those of the mast antenna. A smaller slot antenna with a side feed connection and capacitive load elements was placed into production in the GM 1997 Impact Electric Vehicle.

39.6 FM DIVERSITY ANTENNAS

With the increase in demand for high fidelity FM radio performance, vehicle manufacturers are giving greater attention to reducing the adverse reception effects associated with vehicle movement (i.e., turning corners, traveling in regions of signal blockage, etc.) and multipath signal conditions. These effects can be significantly reduced by using various antenna diversity techniques: space diversity, field-component (polarization) diversity, angle (lobbing antenna pattern) diversity, frequency diversity, and time diversity. However, the general driving public has a negative view of placing several mast antennas on its vehicles. To avoid this porcupine appearance, OEMs generally favor the use of one mast antenna in conjunction with a window glass antenna, or multiple window glass antennas. The outputs of these antennas are processed to obtain best reception using a number of known diversity algorithms such as switch, maximal-ratio, equal-gain, and feed combination.²³ These diversity antennas have a variety of geometric shapes and use multiple locations about the vehicle.²⁴

39.7 RKE AND TPMS ANTENNAS

The remote keyless entry (RKE) and tire pressure monitoring system (TPMS) both operate at the same frequency (315 MHz with a 1-MHz bandwidth) using low RF power levels (to preserve battery life).

RKE Antennas

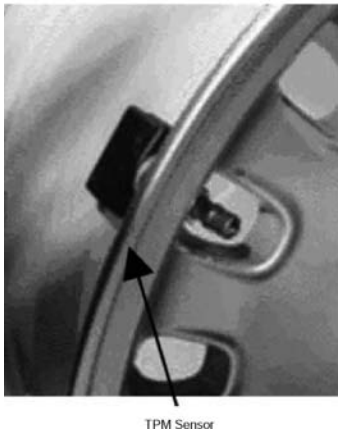
About 80 percent of U.S. cars are sold with RKE systems that consist of a key fob transmitter and a vehicle receiver system for locking/unlocking vehicle doors, unlocking the trunk, and activating an emergency alarm system. The majority of RKE systems use one-way (simplex) communication from an “activated” key fob (i.e., activated by pressing one of the fob buttons) to the vehicle, while more complex systems use two-way communication (half-duplex) to establish an additional channel from the vehicle to the key fob. The two-way system incorporates a “passive” feature where a vehicle’s transmitter is continuously polling to detect the presence of its corresponding key fob. Initially, the “activated” key fob system had a range capability of about 15 m while the “passive” key fob system had a range of about 1 to 2 m. However, the addition of the remote-start function has resulted in an increase in operating range to about 150 m with interest to extending it to 500 m or more. RKE antenna specifications are commonly stated in terms of its operating characteristics. For example, it may be required that the minimum operating distance of the system be measured every 15 degrees about the vehicle with at least 90 percent of these distances being greater than 15 m.

The antenna inside the fob is generally a small printed-circuit loop antenna that couples through the user’s “hand-arm-body,” thereby creating an effective “extension” of the fob’s antenna. The arm and body position of the user has a significant impact on the antenna’s operating range. However, some of the newer longer-distance operating systems are using dielectric fob antennas that greatly reduce this coupling. The RF path loss between the fob transmitter and RKE receiver depends on the distance between them, the height of each, and reflection characteristics of the earth’s surface. The transmitted RF signal travels along a direct path, and a reflected ground path. These two signals will combine or interfere with each other causing the received signal power to fluctuate about its expected inverse square of the distance (“free space”) characteristic for distances less than 10 m and inverse fourth of the distance characteristic for distances greater than 10 m.²⁵ These propagation characteristics must be taken into account in the RKE’s power budget considerations.

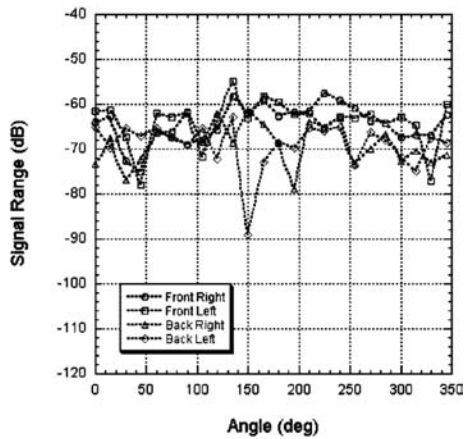
The receiver antenna located inside the vehicle can be mounted internal or external to the receiver box. An internal antenna is generally a wire, meander-line, or loop that is designed to fit inside the plastic housing of the receiver. This arrangement requires the antenna to be fairly small and in close proximity to the receiver’s electronic circuitry (thereby potentially creating self-induced EMC problems that reduce receiver sensitivity). An external antenna can avoid the size restriction and self-induced EMC problems. An external antenna requires the addition of a coaxial cable and an appropriate antenna mounting location. The use of an external antenna significantly increases the operating range of an RKE system over that of an internal antenna. A “hands-free” (passive) RKE system can require up to five antennas (mounted in different locations throughout the interior of the vehicle) to improve azimuth coverage and to determine the fob location (e.g., inside or outside the vehicle).

TPMS Antennas

The function of the tire pressure monitoring systems is to measure the pressure and temperature inside the tires of the vehicle. The majority of these systems use a battery-type



(a) TPM value stem sensor



(b) Normalized simulation results

FIGURE 39-17 A TPM system (after H. J. Song et al²⁶ © IEEE 2004)

microprocessor module to monitor the pressure and temperature of each tire, and an RF transmitter to transmit this data to an in-vehicle receiver. To help extend the microprocessor's battery life (to ten years), a motion sensor is used to control the data transmitting rate (e.g., a longer rate when the vehicle is parked, and a shorter rate when it is in motion). A vehicle with a TPMS can have as many as five transmitting antennas (including one for the spare tire) and one receive antenna. The tire sensor is usually mounted on the drop-center of the wheel or as part of the wheel's valve stem assembly as shown in Figure 39-17a.²⁶ The valve stem antenna is a very small antenna with a length of about $\lambda/100$. Plotted in Figure 39-17b²⁶ is simulated data for a four-wheel TPMS system with a maximum max/min value of 35 dB that was in close agreement with experimental data. TPMS antenna specifications are commonly stated in terms of the system's operating characteristics, and can take the form of sensor operating status versus the angle of rotation for each wheel.

39.8 UHF/VHF TV ANTENNAS

The recent introduction of DVD television systems into the U.S. automobile and SUV market has generated interest in the possible addition of a mobile terrestrial TV system in U.S. vehicles. The cost of these systems is a significant consideration since a good mobile TV system requires a complex diversity receiving system with several vehicle antennas that operate over a very wide frequency bandwidth (i.e., 54–88, 174–216, and 470–806 MHz). This system should really consist of two separate sub-diversity systems: one for picture quality and another for audio quality.²⁷ The quality of reception will improve as the number of antennas is increased, with a minimum of four antennas needed for satisfactory picture and audio reception. The picture and audio sub-diversity systems can use the same diversity antenna system, but each will require its own signal processor module. Similar types of mobile TV systems are common in both Europe and Japan, but to date they have not been widely accepted by the U.S. motoring public.

39.9 MOBILE TELEPHONE ANTENNAS

In 1974, the FCC allocated two 20-MHz bands in the 800 to 900 MHz region for cellular telephone use. The rapid growth in the use of cellular telephones has resulted in both of these bands being increased by an additional 5 MHz. The cellular system was designed to operate as a full-duplex system requiring simultaneous transmission in both the downlink direction (from the tower to the cell phone) and uplink direction (from the cell phone to the tower). To avoid interference between transmit and receive channels, a 45-MHz guard band was allocated between them, as listed in Table 39-2. This cell phone technology was followed by the addition of the Personal Communication System (PCS) in the United States and the Global System for Mobile Communication (GSM), which is the de facto wireless telephone standard in Europe. Wideband Code Division Multiple Access (W-CDMA) is a new 3G global cellular technology that is presently being launched into commercial service.

Dual-band phones operate in two separate frequency bands such as cellular and PCS (for North America), or GSM 900 and GSM 1800 (for Europe, Asia, etc.). Tri-band phones operate in cellular, PCS, and GSM 1800 (for North America, with limited use in Europe, Asia, etc.), or GSM 900, GSM 1800, and PCS (for Europe, Asia, etc., with limited use in North America). Finally, quad-band phones operate in all four of the mobile bands (cellular, GSM 900, PCS, and GSM 1800), permitting their use on a global basis.

Shown in Figure 39-18 are some typical geometries for cell phone mast antennas. A short mast antenna (i.e., height $\leq \lambda/4$) is commonly mounted on the vehicle roof to achieve good gain performance and near omnidirectional radiation/reception patterns in the azimuth plane. However, pattern characteristics of these short antennas are strongly dependent on the sheet metal of the vehicle. A short mast mounted near the center of the roof will have best gain and near omnidirectional patterns. As the mast antenna is moved near the edge of the roof, its gain can decrease and radiation/reception patterns become directional. To help avoid car body effects, a collinear antenna can be used with a $\lambda/2$ vertical antenna element positioned above and connected to a helical-coil positioned above and connected to a $\lambda/4$ mast antenna element. The collinear antenna is generally considered to be independent of a ground plane and is commonly used as an on-glass mounted antenna where its increased height will raise its top element above the roof line of the vehicle, providing an unobstructed azimuth view. A low-impedance adapter is used to couple the RF energy from the antenna through the glass window to the radio coaxial cable.²⁸ Antenna specifications as supplied by automobile manufacturers generally require a gain similar to that of a center roof-mounted $\lambda/4$ mast with omnidirectional radiation/reception patterns (within \pm a few dB) in the azimuth plane. Diversity reception techniques are being investigated to improve mobile telephone reception.²⁹

TABLE 39-2 Land Mobile Telephone Systems

System	Coverage area	Uplink (MHz)	Downlink (MHz)
Cellular	North America	824–849	869–894
GSM 900	Europe	880–915	925–960
GSM 1800	Europe	1710–1785	1805–1880
PCS	North America	1850–1910	1930–1990
WCDMA	World wide	1920–1980	2110–2170

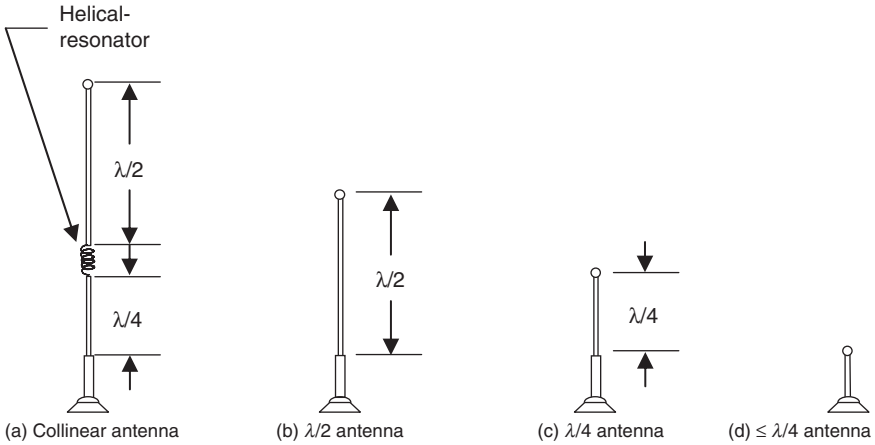


FIGURE 39-18 Typical cellular telephone mast antennas

The requirement by vehicle manufacturers for multiband antenna systems (i.e., with an aesthetically pleasing appearance) has motivated the development of a family of meander-line antennas. These antennas are basically mast antennas designed with a meander-line pattern³⁰ to effectively increase their electrical length to a near resonance condition while limiting their overall physical height to be less than $\lambda/4$ (see Figure 39-19 and Delphi's web site³¹). These types of antennas have been designed using a genetic algorithm technique and will generally require a matching circuit.

Multiband planar inverted F antenna (PIFA) technology has been investigated for vehicle mobile communication purposes. Although these type antennas have been used successfully in mobile telephones and laptop computers, their physical size and on-vehicle directional patterns have not met OEM requirements.

39.10 GPS ANTENNAS

The Global Positioning System (GPS) uses 24 satellites placed in six orbital planes (four satellites in each) that are equally spaced and inclined at a 55-degree angle relative to the equatorial plane of the earth. The altitudes of these satellites are such that they repeat the

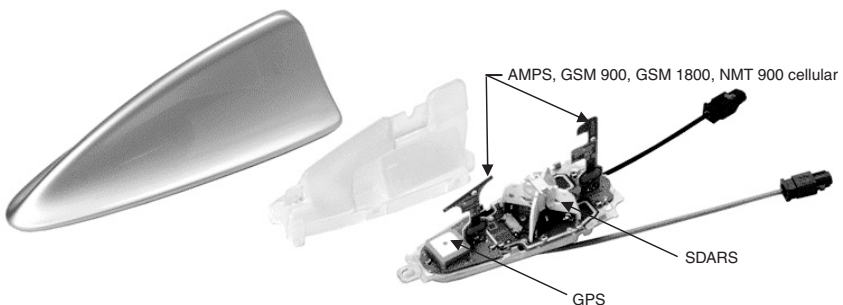


FIGURE 39-19 Multiband vehicle antenna system (Courtesy of Delphi)

same orbit trace over the earth's surface about every 24 hours. This provides a GPS user with potentially five to eight usable satellites from any point on the earth's surface. Each satellite transmits a right-hand circularly polarized (RHCP) signal at a carrier frequency of 1575.42 MHz. This signal is modulated with a unique Pseudo Random Noise (PRN) code within a 1-MHz bandwidth. The PRN code is used to identify which satellite generated a particular GPS signal. These signals are then used to mathematically determine a receiver's position (longitude, latitude, and elevation) by determining its distance from at least four satellites (i.e., using measured time delays) and applying triangulation techniques. The fourth distance is required to determine a precise-time variable that is used in the solution process (x , y , z , and t). The Standard Positioning Service (SPS) is specified to provide accuracies of better than 13 m (95 percent) in the horizontal plane, 22 m (95 percent) in the vertical plane, and 40 ns (95 percent) in time.³² The use of Differential GPS (DGPS) techniques can improve positioning accuracy to within a few meters.

An automobile GPS antenna should be compact, lightweight, reliable, low cost, and easily integrated into the automobile production process. Its beamwidth needs to encompass the greater portion of the hemisphere above the earth's surface (i.e., omnidirectional reception). An example of a GPS antenna pattern is shown in Figure 39-20. The roof is the best mounting location for this type of antenna because it is the highest point on the automobile, providing an unobstructed view of the upper hemisphere, a nearly horizontal mounting surface, a relatively safe site from damage, and easy access for maintenance purposes. The antenna should be kept several inches from the edge of the roof, thereby reducing edge diffraction effects to acceptable levels. If the vehicle has a sunroof, the antenna should be

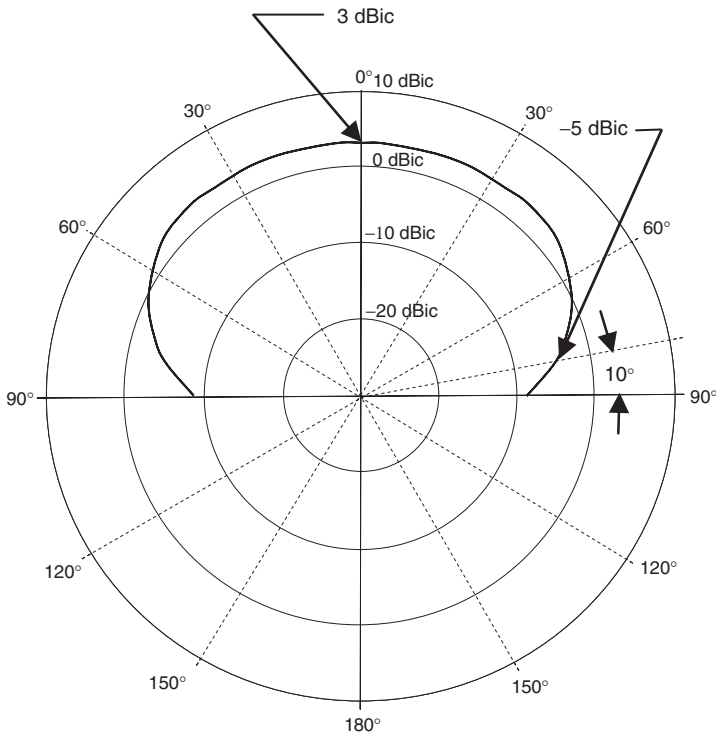


FIGURE 39-20 Example of an elevation pattern for a GPS antenna

kept several inches from the edges of both the roof and sunroof. Other possible mounting locations include the uppermost central region of the windshield, uppermost central region of the rear window, trunk lid, instrument (dash) panel, and rear packaging shelf. However, these mounting locations were determined to be less desirable in that they generally have an adverse impact on the pattern characteristics of the antenna (e.g., signal blockage effects).

The design process for a GPS antenna begins with an “off-vehicle” phase where the proposed antenna is mounted and tested on a flat 1-m-diameter ground plane. This phase is followed by an “on-vehicle” testing phase to validate that the antenna does meet the requirements of the vehicle manufacturer. If the antenna does not meet these requirements, the design process is repeated. Each vehicle manufacturer has developed its own set of on-vehicle antenna requirements. These on-vehicle antenna requirements are considered proprietary information and must be obtained directly from the manufacturer. Generally, these require a minimum passive antenna gain of 2 dBic in the zenith direction ($\theta = 0^\circ$), and -10 dBic for an elevation angle of 10° ($\theta = 80^\circ$). A low-noise amplifier (LNA) is integrated into the base of the GPS antenna to compensate for cable loss and improve the noise figure (NF) at the input port of the receiver. The LNA generally has a gain between 15–35 dB and NF about 1 dB.

Microstrip Patch Antennas

The microstrip patch antenna has become the most popular type of GPS antenna because of its low profile, small footprint, conformability to vehicle surfaces, manufacturability, physical robustness, and low cost. A typical microstrip antenna consists of a very thin conductive patch ($t \ll \lambda_0$), thin low-loss ceramic substrate ($h \ll \lambda_0$, $h \gg t$, and $\epsilon_r \approx 20$), and finite ground plane. Its RF characteristics include a high Q (quality factor), narrow bandwidth, maximum gain in the zenith direction, and near hemispherical patterns. The theory and design principles for this antenna are extensively reported in the literature,³³ and only a brief overview of its operating principles will be presented.

Insight into the operating principles of the microstrip antenna can be achieved by using either a transmission-line model or cavity model. The transmission-line model is used to view the patch as consisting of two orthogonal low-impedance transmission lines with slot apertures at each of their ends. Each orthogonal line has a length of $\lambda_g/2$ (where λ_g is the wavelength in the dielectric material) creating a phase difference of 180° between its paired apertures, thereby achieving maximum gain in the zenith direction. Circular polarization is achieved by setting the phase difference between the two orthogonal transmission lines to be 90° . This phase difference is obtained by selecting the appropriate geometry for the conductive patch, dielectric substrate, and/or feed system. In a similar manner, the cavity model can be used to view the patch as a dielectric-loaded cavity with dominant resonant TM_{010} and TM_{001} modes as shown in Figure 39-21.

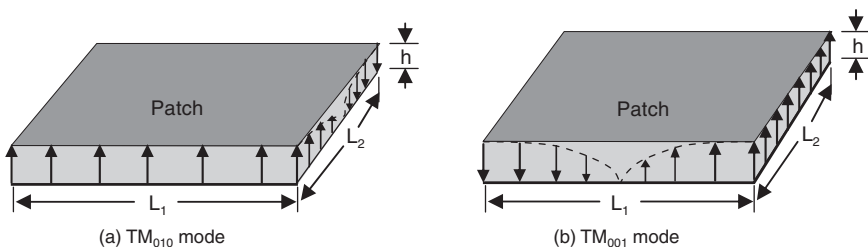


FIGURE 39-21 Mode configurations for a microstrip patch antenna

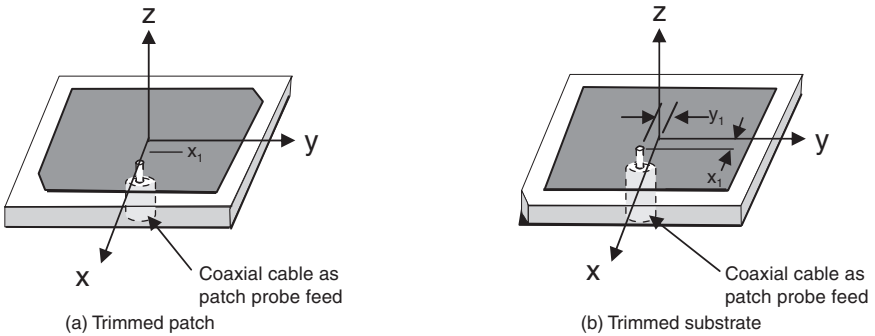


FIGURE 39-22 Typical circularly polarized patch antennas

The bottom-fed coaxial cable connection is a commonly used feed system for patch antennas. Its center conductor is attached to the patch element and its outer shield is attached to the ground plane. An example of a circular polarized patch antenna is shown in Figure 39-22a with a dual trimmed corner patch element and an x_1 offset-feed point. Another example is shown in Figure 39-22b with a single trimmed corner substrate and an x_1 and y_1 offset-feed point.

The design process for the patch element (i.e., its composition, substrate shape, and patch shape) must take into account the electrical characteristics of the vehicle panel on which it is to be mounted (see Figure 39-23 and Reference 34). A sheet metal panel will act as an extended ground plane and affect the RF characteristics of the patch. In addition, edge diffraction effects of the sheet metal panel can have a significant impact on the resulting antenna pattern in regions of low signal levels (i.e., in the form of a ripple pattern being superimposed on the antenna pattern). A patch mounted on a sloped surface will have its patterns tilted in the direction of the sloped surface. These and other antenna-vehicle interactions must be taken into account during the final design and testing phase of the patch antenna.

Shown in Figure 39-24 is a commercially available GPS patch antenna along with its RF characteristics when mounted on a 70-mm-square ground plane.³⁵ The patch design must take into account the dielectric loading of its radome and/or window glass (i.e., if attached to the inside surface of a vehicle window). The patch shown in Figure 39-24 has a 5-MHz frequency offset that allows for the loading effect of its radome. The narrow bandwidth of the patch antenna is commonly used as an effective filter to block RF noise and EMI (electromagnetic interference) signals before they reach the LNA circuit.

Quadrifilar Helix Antennas

The quadrifilar helix (QFH) antenna is recognized as possibly the best GPS antenna, and was the first antenna used for this purpose. The quadrifilar antenna has been investigated for the past several decades, and its characteristics are well known.³⁶⁻³⁸ It has a cardioid-shaped conical pattern, a low front-to-back pattern ratio, and good circular polarization characteristics. It consists of four equally spaced wires that are wrapped around a cylindrical surface, forming four equally spaced helices. A four-way equal signal combiner with relative phases of 0° , 90° , 180° , and 270° is used to connect these helices to the antenna feed port.

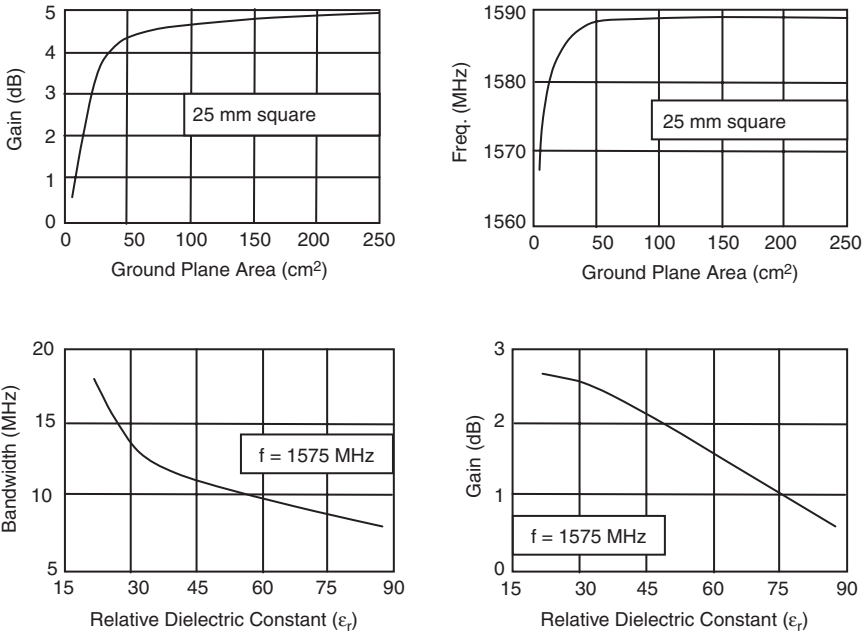


FIGURE 39-23 Relationships between physical and RF properties of a patch antenna (*Courtesy of Spectrum FSY Microwave*)

This antenna can be viewed as two bifilar helix antennas that are physically rotated 90° from each other and fed in phase quadrature. The lengths of its helix elements are an integer multiple of a quarter wavelength ($\lambda/4$) at its resonant frequency. Its low-backlobe characteristic allows it to be operated with or without a ground plane.

The design process for this antenna involves selecting a height, diameter, and pitch angle to obtain desired gain, pattern shape, and impedance. Resonant $\frac{1}{4}$ -turn, $\frac{1}{2}$ -turn, and 1-turn quadrifilar helix antennas generate cardioid-shaped patterns for all axial lengths and diameters. The element ends of the $\frac{1}{4}$ - λ antenna are open-circuited while those of the $\frac{1}{2}$ - λ and 1- λ antennas are short-circuited (see Figure 39-25). Antenna elements with more than one turn and/or small axial length-to-diameter ratios will have degraded pattern shapes and higher polarization axial ratios. Shown in Figure 39-26 is experimental data demonstrating the relationship of the beamwidth, axial ratio, and front-to-back pattern ratio versus axial length.³⁷

Listed in Table 39-3 are the cylindrical dimensions for three $\frac{1}{2}$ -turn, $\frac{1}{2}$ - λ -long antenna geometries designed to operate at 1.575 GHz.³⁸ The process for selecting antenna radius, pitch angle, and height is very time consuming without the use of an EM computer program. The method-of-moments technique was

TABLE 39-3 Fraction Turn Volute Descriptions

	Axial Length (λ)	Volute Radius (cm)
Antenna 1	0.20	1.60
Antenna 2	0.27	1.39
Antenna 3	0.35	1.04

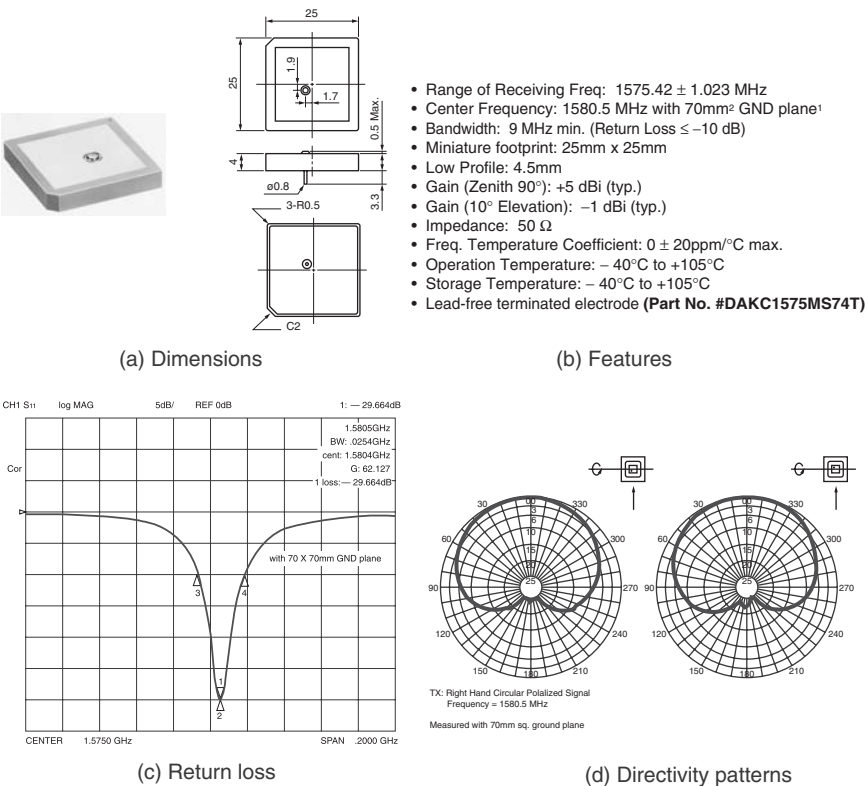


FIGURE 39-24 GPS dielectric patch antenna (Courtesy of Toko America, Inc.)

used to compute the radiation properties for these antennas, and the resulting data is shown in Figure 39-27.

The four-way equal signal combiner with relative phases of 0°, 90°, 180°, and 270° is a critical component for this antenna. One design approach is to use a feed combiner consisting of multiple 90° drop-in hybrid couplers (available in discrete and chip forms). However, the insertion loss for this type of feed is about 0.75 to 1.25 dB,³⁹ which is considered to be unacceptable for both GPS and SDARS applications. Another feed approach that avoids this high insertion loss and has a lower cost is a pair of microstrip power dividers with one feed branch $\lambda/4$ longer (i.e., 90° phase shift) than the other branch. This type of feed system was found to be much more efficient, having an insertion loss of 0.2 to 0.4 dB. Another feed approach is to use two bifilar antennas with one resonant just above and the other just below the desired resonate frequency of

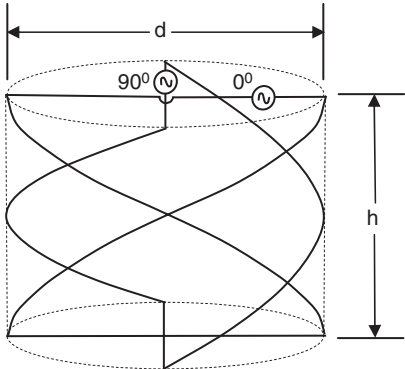


FIGURE 39-25 Quadrifilar helix antenna

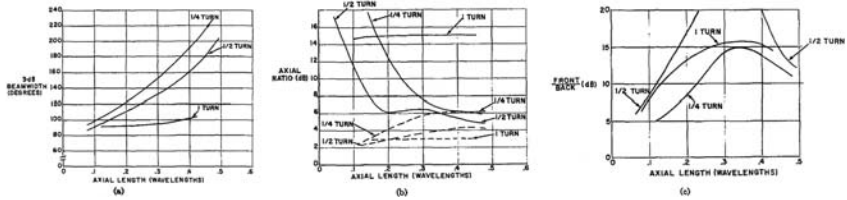


FIGURE 39-26 Quadrifilar helix experimental data: solid lines in (b) indicate peak axial ratio over the entire hemisphere in front of helix; dashed lines indicate peak axial ratio over the 3-dB beamwidth of helix pattern (after C. C. Kilgus³⁷ © IEEE 1969)

the combined quadrifilar helix antenna. These bifilar antennas have equal gains at the desired resonate frequency of the combined quadrifilar helix antenna, but with relative phase shifts of $+45^\circ$ and -45° , respectively. Although the quadrifilar helix antenna is a very good antenna for satellite reception, its cost and size are significant disadvantages.

Dielectric-Loaded Quadrifilar Helix Antennas

The dielectric-loaded quadrifilar helix antenna is an emerging technology consisting of copper tracks deposited onto a high-dielectric ceramic cylinder and individually laser trimmed for optimum antenna performance.⁴⁰ The dielectric cylinder is 18-mm long with a diameter of 10 mm and a relative dielectric constant of about 40. The antenna is formed by two quadrature-phased bifilar helical loops of slightly different lengths to achieve two slightly different offset resonant frequencies (see Figure 39-28). The longer bifilar helix has a lower resonant frequency to create an inductive input impedance with a phase angle of $+45$ degrees at 1.57542 GHz, and the shorter bifilar helix has a higher resonant frequency to create a capacitive input impedance with a phase angle of -45 degrees at the same 1.57542 GHz frequency. These helices are positioned so that they are orthogonal to each other, and fed by a four-way equal signal combiner located at the top of the cylinder. The antenna currents travel around the helices to the rim of an integral sleeve balun that effectively isolates the antenna from the unbalanced coaxial feed cable. This balun technique creates a balanced antenna condition that significantly reduces the near field proximity

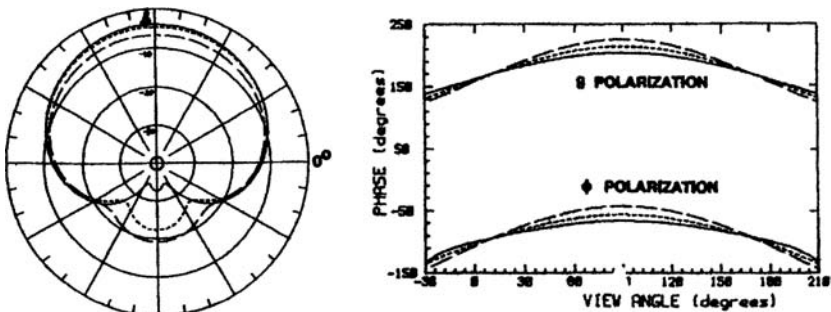


FIGURE 39-27 Computed amplitude and phase response patterns for the $\lambda/2$ quadrifilar helix at 1575.42 MHz with axial length = 0.20λ (—), 0.27λ (---), and 0.35λ (---) (after J. M. Tranquilla and S. R. Best³⁸ © IEEE 1990)

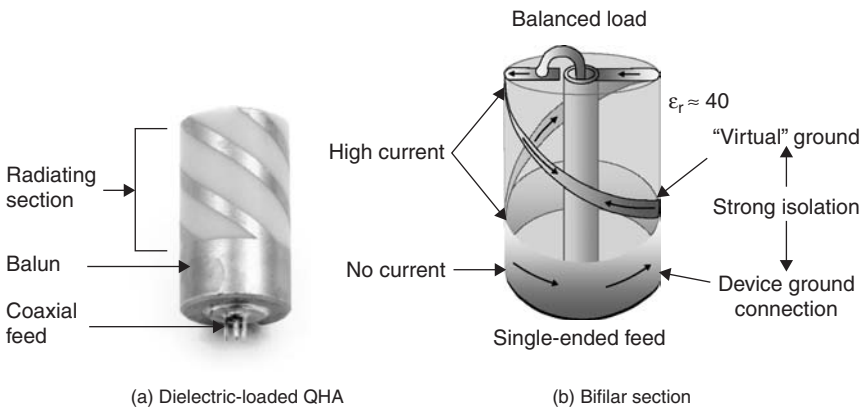


FIGURE 39-28 GPS dielectric quadrifilar helix antenna (Courtesy of Sarantel)

effects between the antenna and GPS receiver. Contained in Figure 39-29 is a list of its specifications and a representative reception pattern. The dielectric-loaded QFH is used in handheld GPS devices and is presently being investigated for vehicle application.

39.11 SDARS ANTENNAS

The Satellite Digital Audio Radio System (SDARS) was placed into service in 2001 to provide its customers with a large number of digital quality radio channels on a nation-wide basis (i.e., for the continental United States). It consists of two service providers (XM Radio, Inc. and Sirius Satellite Radio, Inc.), each with multiple digital radio satellites supported by a network of complementary terrestrial repeaters. SDARS satellites transmit left-hand circularly polarized (LHCP) signals, while terrestrial repeaters transmit vertically polarized signals. Each provider uses different satellite configurations, repeater networks, and antenna requirements (see Table 39-4 and the literature⁴¹⁻⁴³).

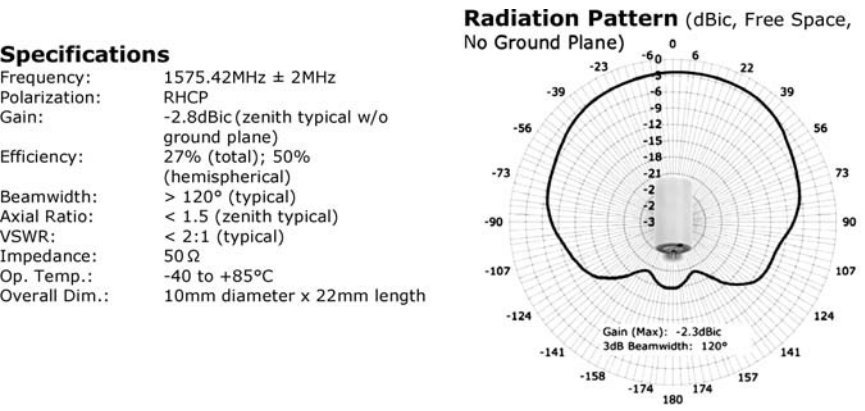


FIGURE 39-29 Specifications and pattern for GPS dielectric quadrifilar antenna (Courtesy of Sarantel)

TABLE 39-4 SDARS Performance Requirements

	Parameter	XM	Sirius
Satellite	Frequency (MHz)	2332.5–2345	2320–2332.5
	Polarization	Left-hand circular	Left-hand circular
	Gain (dBi)*	+2–+4	+2–+4
	Elevation coverage (deg.)	20–60	25 – 90
	LNA gain (dB)	32	36
	LNA noise figure (dB)	0.7–1.2	0.7
Terrestrial	Polarization	Linear	Linear
	Gain (dBi)*	–1	–1
	Elevation coverage (deg.)	0	0
	LNA gain (dB)	30	28
	LNA noise figure (dB)	1.5	2

*For an antenna placed at the center of a 1-m-diameter ground plane

XM has two transmitting geostationary satellites, with each assigned to operate in two of four subfrequency bands about its 2.33875 GHz center frequency (see Figure 39-30). The two transmitting satellites have their orbit tracks located over the equator at 85° and 115° west longitudes with an orbiting radius of 43,000 km. Elevation angles of reception (i.e., in the southerly direction) vary from 60° for the southernmost region of the U.S. to 20° for the northernmost region. Tall buildings, overpasses, tunnels, leafy trees, and other obstacles can significantly “block” these signals. To help mitigate this signal blockage, a network of up to 1500 terrestrial repeaters, operating in two subfrequency bands, are distributed throughout the U.S. (primarily in major cities). These repeaters receive signals from the satellites and retransmit them as vertically polarized signals after appropriate demodulation and remodulation. Satellite signals are modulated using QPSK (Quadrature Phase Shift Keying) while terrestrial signals are modulated using OFDM (Orthogonal Frequency Division Multiplexing). An XM radio can receive signals from three sources: one from each satellite and another from a terrestrial repeater. These signals contain the same program content but travel along different paths. An XM receiver will continuously monitor these signals and use spatial, time, and frequency diversity techniques to obtain best reception.

Sirius has three satellites in an inclined elliptical orbit with a radius varying from 30,300 to 53,500 km. The elevation angle for reception is greater than 57° for at least one of the three satellites. The higher reception angles result in fewer signals being blocked by tall buildings and other obstacles. Each satellite is assigned to operate in one of two subfrequency bands about its 2.32625 GHz center frequency (see Figure 39-30). Only two of the three satellites are active at any one time, to avoid in-band interference. Each satellite will spend approximately 16 hours within its U.S. footprint, with at least one of these satellites always providing service. Once an active satellite is no longer capable of providing nationwide service, it is deactivated and its assigned signal is handed off to another satellite that is reactivated as it enters the U.S. service area. To help mitigate signal blockage, a network of up to 150 terrestrial repeaters, operating in a third subfrequency band, are primarily located in major U.S. cities. A Sirius radio can receive signals from three sources: two from satellites and another from a terrestrial repeater. A Sirius radio continuously monitors these signals, and applies spatial, time, and frequency diversity techniques to obtain best signal reception.

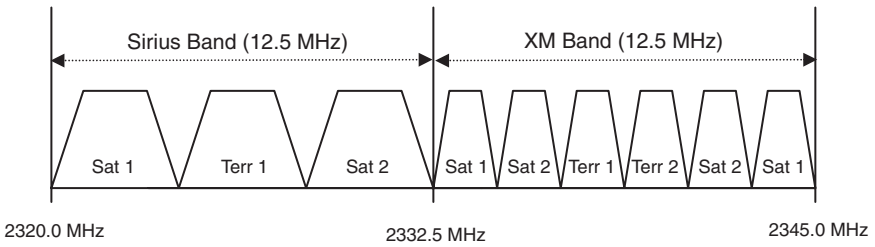


FIGURE 39-30 SDARS frequency allocation

An SDARS automobile antenna is required to have good omnidirectional patterns for both high (celestial) and low (terrestrial) elevation angles. In addition, these antennas must be aesthetically pleasing (i.e., very low profile or preferably hidden), low cost, lightweight, and easily integrated into the vehicle production process. The roof is the best mounting location for the antenna because it is the highest point of the vehicle, providing an unobstructed view of the upper hemisphere, a nearly horizontal mounting surface, a relatively safe site from damage, and easy access for maintenance purposes. The antenna should be located a minimum distance of several inches from the edge of the roof edge in order to reduce diffraction effects to acceptable levels. If the vehicle has a sunroof, the antenna should be kept at a distance of several inches from the edges of the roof and sunroof. Other possible mounting locations include the uppermost central region of the windshield, uppermost central region of the rear window, trunk lid, instrument (dash) panel, and rear packaging shelf. However, these mounting locations were determined to be less desirable in that they generally have an adverse impact on the pattern characteristics of the antenna (e.g., signal blockage effects).

Dual antenna systems (one for celestial reception and another for terrestrial reception) are commonly used to meet the requirements of both the service providers and vehicle manufacturers. The most common satellite antenna is the microstrip patch, and the most common terrestrial antenna is the monopole. Service providers state their antenna requirements in terms of an antenna mounted on a 1-m-diameter ground plane, while vehicle manufacturers state their antenna requirements in terms of an antenna mounted on the intended vehicle. The antenna characteristics for these two mounting conditions can be significantly different. A more detailed discussion of the effects of a finite (1 m diameter) ground plane and its appropriate edge geometry is contained in M. Daginnus et al.⁴⁴ In addition, the low-level signal environment that an SDARS receiver system must operate in requires the use of LNAs (see Table 39-4) to compensate for cable loss and to improve SNR for both satellite and terrestrial reception.

Satellite Antennas

The antenna technologies that were reviewed in Section 39-10 for GPS antennas (i.e., microstrip patch, quadrifilar helix, and dielectric quadrifilar helix antennas) are applicable to SDARS antennas and will not be repeated in this section. As is the case for the GPS antenna, the microstrip patch antenna has become the most popular type of SDARS automobile antenna because of its low profile, small footprint, conformability to vehicle surfaces, manufacturability, physical robustness, and low cost. However, the quadrifilar helix antenna is considered to have better reception characteristics, followed by the bent crossed-dipole antenna.

Bent Crossed-Dipole Antenna

One technique for generating a hemispherical circularly polarized antenna pattern is to place two $\lambda/2$ dipoles above a ground plane and at right angles to each other. These dipoles are fed with signals of equal amplitude but shifted in phase by 90° . The resulting E -plane and H -plane patterns for each dipole have beamwidths (relative to zenith) of 60° and 120° , respectively.⁴⁵ This difference in beamwidths results in low axial ratios for only a few degrees about the zenith. One technique to correct this deficiency is to use bent crossed-dipole antennas that will increase the beamwidth of the E -plane pattern.³⁹ An example of a bent crossed-dipole antenna is shown in Figure 39-31a with its measured and simulated patterns shown in Figure 39-31b. The feed system for this antenna consists of a microstrip power divider, with one arm being $\lambda/4$ longer for the required 90° phase difference.

Terrestrial Antennas

The terrestrial antenna generally takes the form of a simple monopole antenna. In one application, the monopole is positioned along the axis of a quadrifilar helix antenna.⁴¹ In another application, the monopole is located at the center and along the normal axis of a microstrip patch antenna (i.e., extending from the patch base, through an access channel cut into the ceramic material, and out the top). In another application, four electrically short ($\lambda/8$) monopoles are placed symmetrically around the bent crossed-dipole antenna as shown in Figure 39-31. This monopole arrangement forms a terrestrial antenna array, with signals from these monopoles being combined using a microstrip feed assembly. The low height associated with satellite patch antennas allows the use of much smaller terrestrial antennas (e.g., meander-line antenna, dielectric mast antenna, etc.).

PIFAs are commonly used in handheld cell phones⁴⁶ and have been investigated for use as a vehicle SDARS terrestrial antenna. However, their on-vehicle RF characteristics and various physical attributes have significantly limited their potential use as a vehicle antenna.

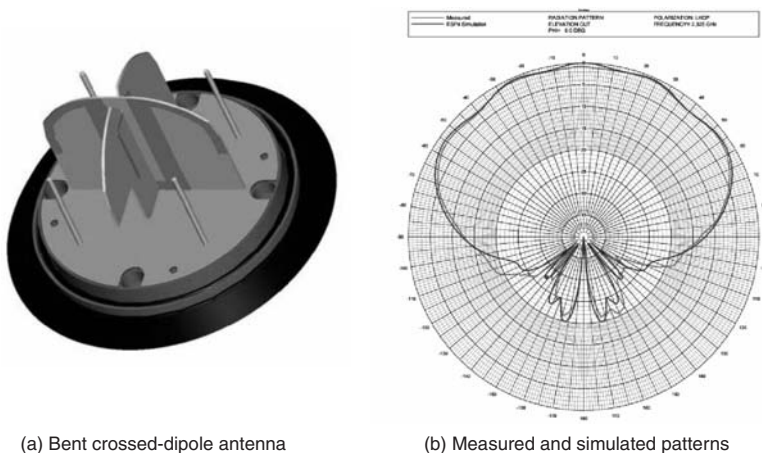


FIGURE 39-31 SDARS bent crossed-dipole antenna with reception patterns (*Courtesy of RecepTec, Inc.*)

The gain for the terrestrial antenna should be greater than -1.0 dBi in the azimuth plane ($\theta \approx 90^\circ$). However, a number OEMs have begun using the patch antenna for both satellite and terrestrial reception. They have determined that the slight loss in terrestrial gain is offset by lower cost and improved aesthetics.

39.12 EMERGING TECHNOLOGIES

Table 39-5 lists a number of emerging automobile telecommunication technologies. Although ZigBee, Bluetooth, and WiFi use the same frequency band and appear to be similar, they are designed for considerably different applications:

- ZigBee for control, monitoring, automation, sensors, etc.
- Bluetooth for mobile phones, headsets, MP3s, PDAs, notebooks, etc.
- WiFi for wireless local area networks (WLANs)

The ZigBee technology is capable of using radio node devices to form mesh networks, where each node can relay information to its neighboring nodes. If a node malfunctions, one or more of its neighboring nodes can reroute communication around the disabled node, providing a robust communication network.

Bluetooth and WiFi technologies use a central hub-and-spoke distribution network that channels communication to its radio devices. Bluetooth offers automatic network configuration and service discovery that makes the establishment of the network almost transparent at distances up to 10 m. WiFi operates much like a mobile phone, providing fast connection from anywhere within its coverage area. It connects computer devices to its network at distances up to 100 m. WiMAX is similar to WiFi but is intended to support a metropolitan-wide wireless network capable of handling thousands of users. It provides high-throughput broadband connections over long distances (with a maximum range of about 30 miles). WiMAX has much promise with its established global standards, high throughput, high reliability, and low-cost services.

TABLE 39-5 Emerging Vehicle Telecommunication Technologies

Application	Frequency Band (GHz)
ZigBee	2.402–2.480
Bluetooth	2.402–2.480
WiFi - IEEE 802.11a	5.150–5.350
WiFi - IEEE 802.11a	5.725–5.825
WiFi - IEEE 802.11b	2.410–2.462
WiFi - IEEE 802.11g	2.410–2.462
WIMAX - IEEE 802.16	2–11
DSRC - IEEE 802.11p	5.850–5.925
Mobile Satellite TV	12.2–12.7
Automobile radar	24 & 76

Dedicated Short Range Communications (DSRC) is a short-to-medium-range wireless system specifically designed for automobile use. This system is being developed to support a wide range of vehicle-to-vehicle and vehicle-to-roadside communications. It will complement the mobile telephone by supporting new communication systems that require time-critical responses and/or high data-transfer rates. It will operate in the 5.850 to 5.925 GHz band and have a range of up to 1000 m. Some proposed applications include electronic toll, electronic parking, warning of an approaching emergency vehicle, cooperative adaptive cruise control, intersection collision avoidance, etc.

Mobile satellite TV is another wireless system being investigated for automobile use. Satellite TV operates in the frequency band from 12.2 to 12.7 GHz (Ku-band) and requires a rather large circularly polarized receiving antenna with a beamwidth of less than 1 degree. The standard TV satellite dish is too big and bulky to be placed on a moving vehicle, and would require some means to keep it pointed at the satellite as the vehicle maneuvers through traffic. The high operating frequency necessitates that the satellite TV antenna be mounted on the roof, or possibly the trunk lid, to avoid signal blockage by the structure of the vehicle. To overcome the dish size limitations, the vehicle antenna should have a "flat" plate geometry with mechanical and/or electronic scanning capability. Proposed mobile TV antennas have not been able to meet the physical size, weight, and cost requirements of vehicle manufacturers. Mobile satellite TV systems are commercially available as an aftermarket or add-on accessory item, but these are currently too heavy, bulky, and costly (e.g., a few inches in height and about 30 inches in width) for common vehicle use.

Automobile radar is another area that is being investigated. Potential applications include collision warning (with autonomous braking), headway distance warning (HDW), adaptive cruise control (ACC) equipped with stop and go, blind spot detection (BSD), pre-crash detection for smart airbags, and parking aid (PA) systems. These radar systems generally operate in the Ka- (e.g., 24 GHz) and W-band (e.g., 76 GHz). Their distance capabilities vary from 2 m up to 150 m with antenna beamwidths from less than ± 4 up to ± 45 degrees. Antenna geometries include a single beam, dual beam, and multiple beams with mono-pulse, mechanical, and electronic scanning capabilities. It has been suggested that automotive radar frequencies may move above 100 GHz in a continuing effort to shrink the physical size of the antenna.

39.13 ANTENNA MEASUREMENTS

Over the past several decades, automobile manufacturers have had to develop their own antenna measurement techniques for characterizing vehicle antennas. This process started with the AM band and was later expanded in the mid-1960s to include FM. In the early 1980s, an R&D automobile antenna range was built to investigate new automobile antenna technologies and determine range requirements for antenna pattern measurements.²¹ A decade later, a much larger and more accurate range was built with the capability of measuring AM, FM, RKE, and cell phone antennas (see Figure 39-32). This quasi-far-field facility was designed to measure vehicle antennas in an RF environment similar to fringe reception conditions. At about the same time, a number of other automobile manufacturers and antenna suppliers built in-house ranges using their own range requirements. Over the following years, it was determined that there are significant discrepancies in the data measured by these ranges, with no consensus as to what data set is correct. In an effort to resolve these discrepancies, an industry/government effort was established to conduct an in-depth RF evaluation of the large outdoor range shown in Figure 39-32, and make appropriate recommendations to improve the accuracy and reduce the uncertainties of its measured data. This study revealed that the in-ground turntable had a significant adverse impact on the uniformity of the electromagnetic

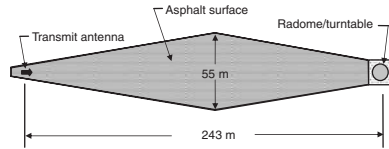


FIGURE 39-32 Far-field antenna facility (*Courtesy of General Motors*)

field just above the turntable surface. Recommendations were made for developing new measurement procedures to improve accuracy, enhance repeatability, and reduce uncertainties of its measured data to within stated error limits.^{47,48}

The introduction of automobile satellite antennas has generated a need for a different type of antenna range capable for measuring vehicle antenna patterns that encompass the greater portion of the hemisphere above the earth's surface (see Figures 39-29 and 39-31). It was determined that a hemispherical near-field antenna test facility was required to make this type of antenna measurement. A near-field system samples the amplitude and phase of the electromagnetic field along a predetermined surface very close to and surrounding the antenna under test. This near-field data is then used to determine its corresponding far-field patterns. For an automobile satellite antenna, the sampling surface is a hemisphere with a radius that is sufficiently large to encompass the entire vehicle. Two basic near-field systems are currently available: an outdoor near-field range (see Figures 39-33 and 39-34 for two examples of this type of system), and an in-door (anechoic chamber) near-field range (see Figure 39-35).

The outdoor near-field range generally consists of an in-ground turntable, a radome, and a single-probe "gantry arm" near-field system.⁴⁹ This near-field system has a wide operating bandwidth (FM band up to the mm band), but its single-probe setup requires several hours to complete measurements for a GPS or SDARS antenna.



(a) Far-field range



(b) Near-field range

FIGURE 39-33 Outdoor near-field antenna facility with radome (*Courtesy of Nippon Antenna Co., Ltd.*)

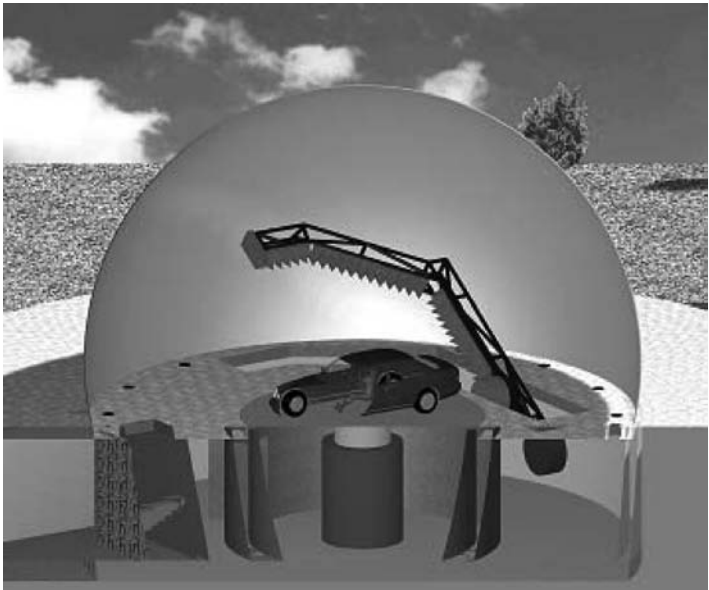
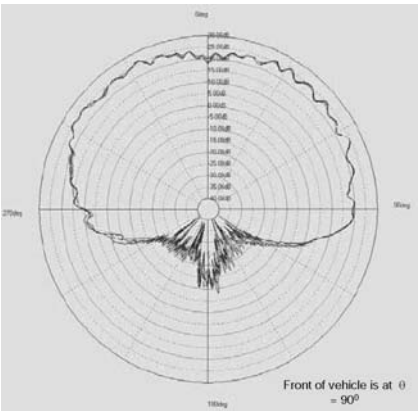


FIGURE 39-34 Outdoor near-field antenna facility with radome (Courtesy of Hirschmann Car Communication, Inc.)

The indoor near-field range consists of an anechoic chamber, a turntable, and a multiprobe arc near-field system.⁵⁰ The multiprobe arc has a 6-meter radius with 103 dual-polarized probes located every degree along its length. The system (including the dual-polarized probes) has an operating bandwidth from 400 MHz to 6 GHz, and is capable of multiple-frequency and/or antenna operation. Its normalized pattern directivity ($D_n = |D(\theta, \phi)/D_{\max}|$) error is less than



(a) Near-field range



(b) Multifrequency elevation patterns for a rear roof-mounted SDARS antenna

FIGURE 39-35 Anechoic chamber near-field antenna facility (Courtesy of General Motors)

± 0.25 dB for relative pattern levels from 0 to -5 dB, and less than $\pm 0.1 \times (-D_n)$ dB for pattern levels from -5 to -30 dB. Peak gain measurements are repeatable to within a few tenths of a dB, and inter-laboratory data comparisons (with an outside RF laboratory) demonstrated a correlation to within 1.0 dB. Major advantages of this facility are its accuracy, repeatability, and speed. The multiple dual-polarized probes allow the system to complete pattern measurements in 5 minutes (limited by the rotational speed of the turntable) instead of the several hours require for the single-probe system.

REFERENCES

1. H. A. Wheeler, "Small Antennas," *IEEE Trans. on Antennas and Propagat.*, vol. AP-23, no. 4 (July 1975): 462-469.
2. A. R. Von Hippel, *Dielectric Materials and Applications* (Cambridge, MA: The MIT Press, 1954): 314.
3. R. A. Burberry, *VHF and UHF Antennas*, IEE Electromagnetic Wave Series 35 (London: Peter Peregrinus Ltd., 1992): 199-209.
4. W. M. Heina, Portable Radio Apparatus (April 26, 1927): U.S. Pat. 1,626,464.
5. C. W. Rowan and C. A. Altgelt, "When Car Radios Were Illegal—A History of Early American and European Car Radios," SAE International Congress and Exposition, Detroit, MI, SAE Paper 850019 (February 1985): 9.
6. E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd Ed. (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1968): 323-326.
7. R. C. Johnson and H. Jasik, *Antenna Applications Reference Guide* (New York: McGraw-Hill, 1987): 1-7.
8. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design* (New York: John Wiley & Sons, Inc., 1981): 47-53.
9. N. Adams, *The Complete Corvette Restoration & Technical Guild – Vol. 1 1953 through 1962* (Princeton, NJ: Princeton Publishing Inc., 1980): 93-94.
10. B. F. Elving, *FM ATLAS and Station Directory*, 12th Ed. (Esco, MN: FM Atlas Publishing, 1989).
11. W. K. Jensen, Concealed Windshield Broadband Antenna (April 27, 1971): U.S. Pat. 3,576,576.
12. H. K. Lindenmeier, J. F. Hopf, and L. M. Reiter, "Active AM-FM Windshield Antenna with Equivalent Performance to the Whip Now as Standard Equipment in Car Production," *IEEE Antennas and Propagation Society International Symposium*, vol. AP-23 (June 1985): 621-624.
13. H. E. Shaw, Combination of Electrically Heated Transparent Window and Antenna (December 16, 1969): U.S. Pat. 3,484,584.
14. L. L. Nagy and J. A. Graham, Broadband FM Vehicle Rear Window Antenna not Requiring a Boost Amplifier (February 29, 2000): U.S. Pat. 6,031,500.
15. L. L. Nagy, D. C. Martin, A. E. Zaczor, and J. A. Graham, Broadcast Radio Signal Seek Circuit (March 30, 2004): U.S. Pat. 6,714,771.
16. J. D. Last, B. Easter, and K. Duffy, "A Broad-Band Heated-Backlite Antenna for a Production Vehicle," SAE International Congress and Exposition, SAE Paper 840278, Detroit, MI, 1984.
17. I. Zafar, Backlite Antenna for AM/FM Automobile Radio (August 4, 1998): U.S. Pat. 5,790,079.
18. L. L. Nagy, M. Bellinger, and D. C. Martin, Transparent Vehicle Window Antenna (June 18, 1996) U.S. Pat. 5,528,314.
19. J. J. Finley, Low Emissivity Film for Automotive Heat Load Reduction (February 6, 1990): U.S. Pat. 4,898,789.
20. L. L. Nagy, D. C. Martin, and M. J. Lewis, Transparent Vehicle Window Antenna with Capacitive Connection Apparatus (February 1, 2000): U.S. Pat. 6,020,855.
21. L. L. Nagy, "A New Generation of Antennas for Automobile Use," SAE International Congress and Exposition, SAE Paper 870092, Detroit, MI, February 23-27, 1987.

22. D. Kightlinger, "The Signal Is Clear: Hidden Antennas Are the Next Wave," R&D Center - Search, General Motors Information Science Department, Warren, MI, vol. 31, no. 1 (July 1997).
23. W. C. Y. Lee, *Mobile Communication Engineering*, Chapters 9 and 10 (New York: McGraw-Hill, 1982).
24. H. K. Lindenmeier and L. M. Reiter, "Multiple-Antenna-Diversity for FM Car Radio Reception," *IEEE Antennas and Propagation Society International Symposium*, vol. 25 (June 1987): 456–459.
25. L. Burgess, "Estimating RKE System Performance," *Microwaves & RF*, <http://www.mwrf.com/Articles/Index.cfm?Ad=1&ArticleID=12041>, February 2006.
26. H. J. Song, H. P. Hsu, R. Wiese, and T. Talty, "Modelling Signal Strength Range of TPMS in Automobiles," *IEEE Antennas and Propagation Society International Symposium*, vol. 3, session 110 (June 20–25, 2004): 3167–3170.
27. H. K. Lindenmeier, L. M. Reiter, R. P. Kronberger, J. F. Hopf, and B. M. Leinwetter, "Terrestrial Television-Reception in Fast Moving Cars by Means of Antenna-Diversity-Techniques," SAE International Congress and Exposition, SAE Paper 1999-01-0259, Detroit, MI, March 1–4, 1999.
28. L. J. Larsen and D. M. Phemister, Mobile Antenna and Through-the-Glass Impedance Matched Feed System (August 16, 1988): U.S. Pat. 4,764,773.
29. H. K. Lindenmeier, L. M. Reiter, J. F. Hopf, and R. P. Kronberger, "Cellular Radio Car Array for Smart Antenna Transmission and Diversity Reception," SAE International Congress and Exposition, SAE Paper 2000-01-0071, Detroit, MI, March 6–9, 2000.
30. K. Fujimoto and J. R. James, *Mobile Antenna Systems Handbook*, 2nd Ed. (Norwood, MA: Artech House, Inc., 2001): 434–436.
31. Delphi Fuba™ Multi Antenna Modules, Delphi Corporation, <http://delphi.wieck.com/imageSingle/ImageID=DEL2004122054561/tpl2=images/type=jpg>.
32. E. D. Kaplan and C. J. Hegarty, *Understanding GPS Principles and Applications*, 2nd Ed. (Norwood, MA: Artech House, Inc., 2006): 1–4.
33. R. E. Munson, *Antennas Engineering Handbook*, 3rd Ed., Chap. 7, R. C. Johnson (ed.) (New York: McGraw-Hill, 1993).
34. "Antenna Element Solutions for Mobile, Portable & Handheld Devices," Spectrum FSY Microwave, www.fsymicrowave.com/pdfs/patchantennabrochure.pdf.
35. "DAKC1575MS74T, GPS Dielectric Patch Antenna," TOKO America, Inc., <http://www.tokoam.com/passives/antennas/pdf/DAKC1575MS74T.pdf>.
36. C. C. Kilgus, "Multielement, Fractional Turn Helices," *IEEE Trans. on Antennas and Propagat.*, vol. AP-16, no. 4 (July 1968): 499–500.
37. C. C. Kilgus, "Resonant Quadrifilar Helix," *IEEE Trans. on Antennas and Propagat.*, vol. AP-17, no. 3 (May 1969): 349–351.
38. J. M. Tranquilla and S. R. Best, "A Study of the Quadrifilar Helix Antenna for Global Positioning System (GPS) Applications," *IEEE Trans. on Antennas and Propagat.*, vol. AP-38, no. 10 (October 1990): 1545–1550.
39. R. Marino and A. Fuchs, "Dual Antenna System for Mobile SDARS Application," SAE 2001 World Congress, SAE Paper 2001-01-0008, Detroit, MI, March 5–8, 2001.
40. "GeoHelix-P2 Passive GPS Antenna," Sarantel, http://www.sarantel.com/downloads/specifications/geohelix-p2_v2.pdf.
41. A. Petros, I. Zafar, and S. Licul, "Reviewing SDARS Antenna Requirements," *Microwave & RF Journal* (September 2003): 51–62.
42. S. J. Patsiakos, "XM Satellite Radio Technology Fundamentals," SAE 2001 World Congress, SAE Paper 2001-01-1328, Detroit, MI, March 5–8, 2001.
43. A. Petros, I. Zafar, and R. Pla, "Antenna Measurement Techniques for SDARS Antennas," Antenna Measurement Techniques Association, Proceedings AMTA-2004, Session 5, PID058, Stone Mountain Park, GA., October 17–22, 2004.

44. M. Daginnus, R. Kronberger, A. Stephan, G.-H. Hassmann, H. Lindenmeier, J. Hopf, and L. Reiter, "SDARS – Antennas: Environmental Influences, Measurement, Vehicle Application Investigations and Field Experiences," SAE 2002 World Congress, SAE Paper 2002-01-0120, Detroit, MI, March 4–7, 2002.
45. R. H. DuHamel and A. R. Mahnad, *Antenna Engineering Handbook*, 3rd Ed., R. C. Johnson (ed.) (New York: McGraw-Hill, Inc, 1993): 28-8 to 28-11.
46. K. L. Wong, *Planar Antennas for Wireless Communications* (New York: John Wiley & Sons, Inc., 2003).
47. R. Johnk, D. Novotny, C. Grosvenor, N. Canales, J. Graham, D. Martin, S. Yencer, D. Hibbard, T. Roach, and L. Nagy, "An Electric-Field Uniformity Study of an Outdoor Vehicular Test Range," IEEE EMC Symposium, August 14–18, 2006.
48. T. L. Roach and L. L. Nagy, "EM Simulation Analysis of an Antenna Range," Delphi Research Labs, Shebly, MI, Delphi Research Labs Report DRLR-0289, March 2005.
49. P. N. Betjes, D. Pototzki, and D. J. van Rensburg, "A Hemi-Spherical Near-Field System for Automotive Antenna Testing," AMTA-2005, Section 15, PID-316, Newport, RI, October 30–November 4, 2005.
50. J. A. Graham and P. O. Iversen, "Rapid Spherical Near-Field Antenna Test System for Vehicle Mounted Antennas," Antenna Measurement Techniques Association, Proceedings AMTA-2004, Section 02, PID-130, Stone Mountain Park, GA, October 17–22, 2004.

Chapter 40

Aircraft Antennas

Gerald J. Oortman

*Senior Electromagnetics Engineer
(Lockheed Martin Aeronautical Systems Company)*

Charles E. Ryan, Jr.

*Consultant
(Formerly, Georgia Institute of Technology)*

CONTENTS

40.1	INTRODUCTION	40-2
40.2	LOW-FREQUENCY ANTENNAS	40-3
40.3	HIGH-FREQUENCY COMMUNICATIONS ANTENNAS	40-9
40.4	UNIDIRECTIONAL VERY-HIGH-FREQUENCY ANTENNAS	40-13
40.5	OMNIDIRECTIONAL VERY-HIGH-FREQUENCY AND ULTRAHIGH-FREQUENCY ANTENNAS	40-14
40.6	OMNIDIRECTIONAL L-BAND FREQUENCY ANTENNAS	40-18
40.7	GPS ANTENNAS	40-20
40.8	ALTIMETER ANTENNAS	40-20
40.9	ANTENNA PERFORMANCE AND AIRCRAFT INTERACTION	40-20

40.1 INTRODUCTION

The design of antennas used for aircraft differs from other applications. Special consideration must be given to the environment to which the antenna is exposed, interaction of the aircraft with the antenna performance characteristics, safety-of-flight, and regulatory requirements.

Aircraft antennas must be designed to withstand severe static and dynamic stresses. Antennas for high speed aircraft must be designed aerodynamically, to limit physical stress on the antenna and limit degradation of the aircraft's flight characteristics. Triboelectric charging of the airframe surfaces by dust or precipitation particles, known as p static, gives rise to corona discharges. These discharges may produce extreme electrical noise, especially when the location of the antenna is such that there is strong electromagnetic coupling to the discharge point. In transmitting applications, corona discharge from the antenna element may limit the power-handling capacity of the antenna. Also, special consideration must frequently be given to the protection of aircraft antennas from damage due to lightning strikes.

The size and shape of the airframe play a major role in determining the impedance and radiation pattern characteristics of an antenna. The type of antenna used in a given application will often depend on the size of the airframe relative to the wavelength. In the case of propeller-driven aircraft and helicopters, the motion of the blades may give rise to modulation of the radiated signal. Through careful placement of the antenna on the aircraft, optimum electrical performance characteristics can be achieved.

While the types of antennas and their locations on an aircraft are influenced by the previous factors, the most important consideration is safety-of-flight. Structural failure of the antenna cannot significantly endanger personnel aboard the aircraft or critical aircraft components (such as propellers, jet engines, wings, flaps, and so on). Redundant systems are installed to ensure the aircraft has at least a minimum level of functionality for safe flight. The type and number of systems required include voice and identification communications for air traffic control as dictated by national and international governments and organizations. The following table shows an example of the types of the avionics systems required for large aircraft.

System Description	Quantity	Frequency	Polarization	Function
ADF	1–2	190–1750 KHz	Vertical	Automatic Direction Finding
HF Voice	1–2	3–25 MHz	Horizontal	Long Range Communication
VHF Voice (FM)	1	30–76 MHz	Vertical	Communication with Air Traffic Control
Marker Beacon	1	75 MHz	Horizontal	Instrument Landing System
VOR/Localizer	2	108–118 MHz	Horizontal	Bearing/Instrument Landing System
VHF Voice (AM)	2	118–152 MHz	Vertical	Civil Aviation
UHF Voice	2	224–400 MHz	Vertical	Military
Glide Path	1	329–335 MHz	Horizontal	Instrument Landing System
DME/TACAN	2	962–1214 MHz	Vertical	Navigation Distance and Bearing
ATC Transponder/IFF	2	1030/1090 MHz	Vertical	Traffic Control/Identification
GPS	1	1525–1585 MHz	RH Circular	Navigation
Radar Altimeter	1	4.2–4.4 GHz	Horizontal	Navigation/Ground Proximity
Weather Radar	1	9.375 GHz	Horizontal	Navigation/Weather Avoidance

40.2 LOW-FREQUENCY ANTENNAS

The wavelengths of frequencies below about 2 MHz are considerably larger than the maximum dimensions of most aircraft. Because of the inherently low radiation efficiency of electrically small antennas and the correspondingly high radio-frequency (RF) voltages required to radiate significant amounts of power, nearly all aircraft radio systems operating at these lower frequencies are designed so that only receiving equipment is required in the aircraft.

Radiation patterns of aircraft antennas in this frequency range are simple electric or magnetic dipole patterns, depending upon whether the antenna element is a monopole or a loop. The pattern produced by a small monopole antenna will always be that of a simple dipole regardless of location; the orientation of the equivalent-dipole axis with respect to the vertical will depend upon location. A small antenna element placed on the airframe would respond to vertical, longitudinal, and transverse field components, indicating that the dipole moment of the antenna-airframe combination has projections in all three directions. Since the aircraft surface is small in terms of square wavelengths, it is appropriate to employ the method of moments (MOM) to assess pattern performance.¹⁻³ This technique is discussed in Section 40.9.

The sensitivity of low-frequency (LF) antennas is customarily expressed in either of two ways, depending upon whether the antenna is located on a relatively flat portion of the airframe such as the top or bottom of the fuselage or at a sharp extremity such as the tip of the vertical stabilizer. In the first case, it can be assumed (at least when the antenna element is small relative to the surface radii of curvature) that the antenna performs as it would on a flat ground plane, except that the incident-field intensity which excites it is greater than the free-space incident-field intensity because of the field fringing produced by the airframe. The effect of the airframe on antenna sensitivity can be expressed by the ratio of the local-field intensity on the airframe surface to the free-space incident-field intensity. For a vertically polarized incident field (which is the case of primary importance in LF receiving-antenna design), this ratio is designated as F_v and is called the *curvature factor for vertical polarization*. Measurements can be performed to determine the equivalent dipole tilt angle and the factor F_v .

Effective height and capacitance data for antennas installed in locations for which this method is applicable are usually obtained by measurements or calculations for the antenna on a flat ground plane. The effective height (for vertically polarized signals) of the antenna installed on the airframe is then estimated by multiplying the flat-ground-plane effective height by the factor F_v appropriate to the installation. On the other hand, the capacitance may be assumed to be the same as that determined with the antenna on a flat ground plane. The presence of a fixed-wire antenna on the aircraft may have a significant effect on F_v . Because of the shielding effect due to a wire, LF antennas are seldom located on the top of the fuselage in aircraft which carry fixed-wire antennas.

Flat-ground-plane data for a T antenna are shown in Figure 40-1. The capacitance curves shown apply to an antenna made with standard polyethylene-coated wire. The standard is a 0.052-in- (1.32-mm-) diameter conductor and a 0.178-in- (4.52-mm-) diameter polyethylene sheath. Two antennas of this type are frequently located in close proximity on aircraft having dual automatic-direction-finder (ADF) installations. The effects of a grounded T antenna on h_e and C_a of a similar nearby antenna are shown in Figure 40-2. To determine the significance of capacitance interaction it is necessary to consider the Q 's of the input circuits to which the antennas are connected and the proximity of the receiver frequencies.

Sensitivity data for a flush antenna and a low-silhouette antenna consisting of a relatively large top-loading element and a short downlead are shown in Figures 40-3 and 40-4, respectively.

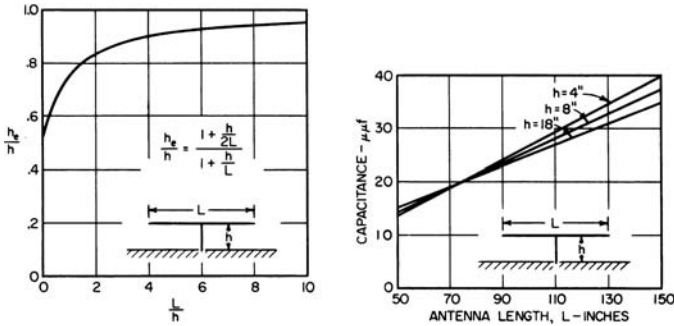


FIGURE 40-1 Effective height and capacitance of a T antenna

The antenna dimensions shown in these figures are not indicative of antenna sizes actually in use, but are rather the sizes of the models used in measuring the data. Both h_e and C_a scale linearly with the antenna dimensions.

A rule-of-thumb design limit for flush antennas may be derived on the basis of a quasi-static analysis of LF antenna performance. If any electric dipole antenna is short-circuited at its feed terminals and placed in a uniform electrostatic field with its equivalent-dipole axis aligned parallel to the field, charges of $+q$ and $-q$ will be induced on the two elements of the antenna. It may be shown that the product of the LF parameters h_e and C_a is related to the induced charge by the equation

$$h_e C_a = q/E \quad (40-1)$$

where h_e and C_a are expressed in meters and picofarads, respectively, q is expressed in pico-coulombs, and the incident-field intensity E is expressed in volts per meter. The quantity q/E is readily calculated for a flush cavity-backed antenna of the type shown in Figure 40-3, at least for the case in which the antenna element virtually fills the cutout in the ground plane. In this case, the shorted antenna element will cause practically no distortion of the normally incident field, and the number of field lines terminating on the element will be the same as the number which would terminate on this area if the ground plane were continuous. The value of q/E is hence equal to $\epsilon_0 a$, where $\epsilon_0 = 8.85 \times 10^{-12}$ F/m and a is the area of

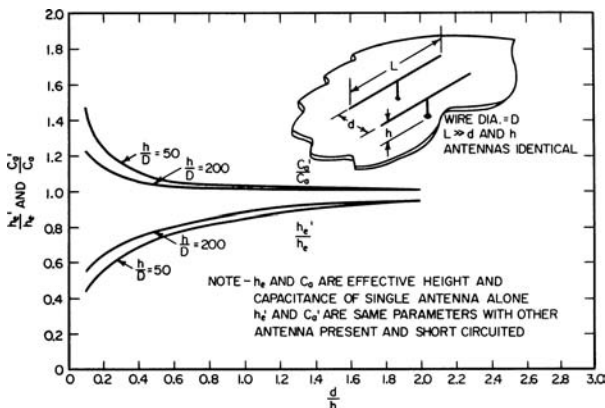


FIGURE 40-2 Shielding effect of one T antenna on another

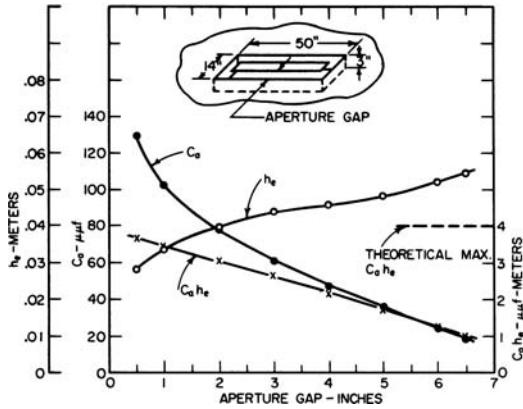


FIGURE 40-3 Design data for a flush low-frequency antenna

the antenna aperture in square meters. For a flush antenna with an element that nearly fills the antenna aperture the result is

$$h_e C_a = \epsilon_0 a \quad (40-2)$$

The line labeled “theoretical maximum $C_a h_e$ ” in Figure 40-3 was calculated from Eq. 40-2. With practical antenna designs, there will be regions of the aperture not covered by the antenna element, so that some of the incident-field lines will penetrate the aperture and terminate inside the cavity. As a result, the induced charge q will be smaller than that calculated previously, and the product $h_e C_a$ will be smaller than the value estimated from Eq. 40-2.

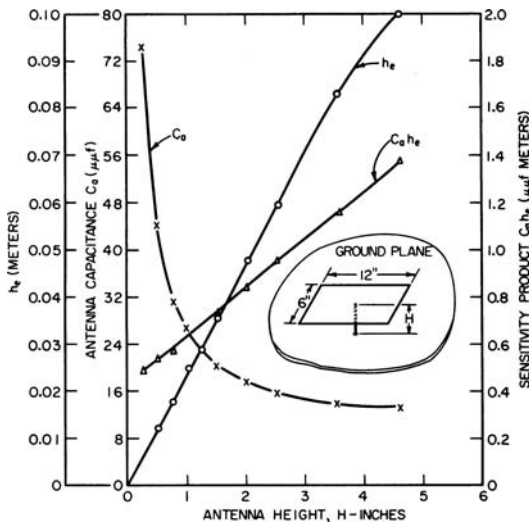


FIGURE 40-4 Design data for a heavily top-loaded low-frequency antenna with flat-plate top loading

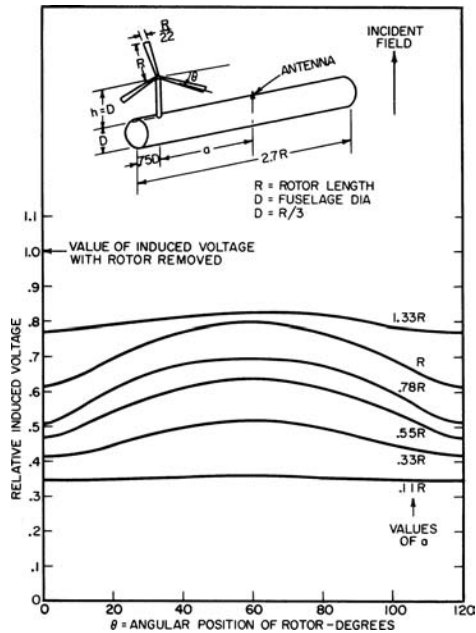


FIGURE 40-5 Rotor modulation for a low-frequency receiving antenna on an idealized helicopter model

The helicopter presents a special antenna-design problem in the LF range as well as in other frequency ranges because of the shielding and modulation caused by the rotor blades. Measured curves of F_v , obtained with an idealized helicopter model to demonstrate the effects of different antenna locations along the top of the simulated fuselage, are shown in Figure 40-5. The MOM also may be used to compute rotor modulation effects. See Section 40.9 for MOM details.

It is characteristic of the rotor modulation of LF signals that relatively few modulation components of significant amplitude are produced above the fundamental rotor modulation frequency (which is equal to the number of blade passages over the antenna per second). This fundamental frequency is of the order of 15 Hz for typical, three-blade, single-rotor helicopters. It should be noted that new sidebands will be generated about each signal sideband and about the carrier. It is thus not a simple matter to predict the effects of rotor modulation on a particular system unless tests have been made to determine the performance degradation of the airborne receiver due to these extra modulation components.

ADF Antennas

The ADF system, which determines the direction of arrival of the signal by rotating its loop antenna about a vertical axis until a null is observed in the loop response (Figure 40-6), is subject to bearing errors because of the difference in direction of the local

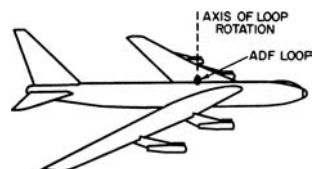


FIGURE 40-6 ADF loop antenna on aircraft

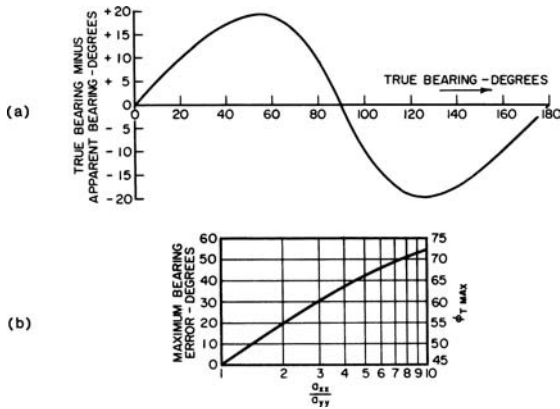


FIGURE 40-7 (a) Bearing-error curve for $a_{zz}/a_{yy} = 2$; (b) The maximum bearing error and the true bearing at which the maximum bearing error occurs as functions of a_{zz}/a_{yy}

incident magnetic fields. The relationship between the true and the apparent directions of the signal source is given by the equation*

$$\tan \phi_t = \frac{a_{xx}}{a_{yy}} \tan \phi_a \quad (40-3)$$

where ϕ_t and ϕ_a are respectively the true and the apparent bearings of the signal source. A curve of the bearing error ($\phi_t - \phi_a$) as a function of ϕ_t with the ratio $a_{xx}/a_{yy} = 2$ is shown in Figure 40-7a. Figure 40-7b shows graphs of the maximum bearing error $|\phi_t - \phi_a|$ and the value of ϕ_t (in the first quadrant) at which the maximum bearing error occurs as functions of the ratio a_{xx}/a_{yy} .

Survey data showing the coefficient a_{xx} and the maximum bearing error for loop antennas along the top and bottom centerline of a DC-4 aircraft are shown in Figure 40-8. The ADF loop has a 180° uncertainty in null location due to its figure eight pattern. An auxiliary sense antenna is used to resolve the null uncertainty via amplitude comparison. The ADF system is calibrated after installation to achieve proper performance.

The automatic-direction-finder loop antenna is normally supplied as a component of the ADF system, and the problem is to find a suitable airframe location. The two restrictions which govern the selection of a location are (1) that the cable between the loop antenna and receiver be of fixed

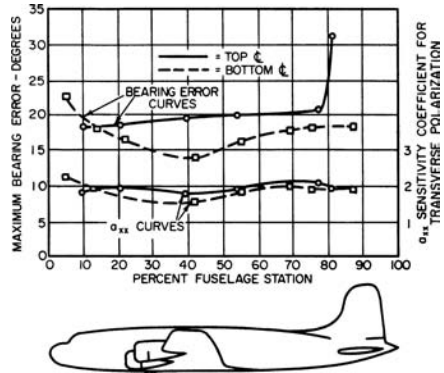


FIGURE 40-8 Low-frequency loop-antenna survey data for a DC-4 aircraft, showing maximum bearing error and sensitivity coefficient for transverse polarization for top and bottom centerline locations

* The ratio of local magnetic field intensity on the airframe surface to the incident magnetic field intensity is designated as a_{xx} for the case in which the incident field is transverse to the line of flight (the signal arrives from the front or rear of the aircraft). It is designated as a_{yy} when the incident field is along the line of flight (the signal arrive from the side of the aircraft).

length, since the loop and cable inductances form a part of the resonant circuit in the loop amplifier input, and (2) that the maximum bearing error which the system can compensate be limited to about 20° . As can be seen from the bearing-error data in Figure 40-7, positions on the bottom of the fuselage forward of the wing meet the minimum bearing-error requirements on the aircraft type for which the data are applicable (and on other comparable aircraft). Such positions are consistent with cable-length restrictions in most cases since the radio-equipment racks are usually just aft of the flight deck.

Bearing-error compensation is accomplished in some ADF systems by means of a mechanical compensating cam.⁴⁻⁵ Standard flight-test procedures⁴⁻⁶ are generally used in setting the compensating cams. In some loop designs⁷ electrical compensation is achieved by modifying the immediate environment of the actual loop element.

Unlike the loop antenna, the sense antenna must be located within a limited region of the airframe for which the equivalent-dipole axis is essentially vertical. This ensures accurate ADF performance as the aircraft passes over or near the ground station to which the receiver is tuned.⁸

A typical flush ADF sense-antenna installation is illustrated by the C-141 aircraft.⁹ The antenna, in this case, is a cavity built into the upper fuselage, fairing just aft of the wing crossover. It was decided during the design phase to use the Bendix ADF-73 automatic-direction-finder system, which to some extent determined the sense-antenna requirements. The basic antenna requirements were (1) antenna capacitance of 300 pF, (2) a minimum quality factor of 1.4, and (3) a radiation-pattern tilt angle between 0 and 10° downward and forward. Figure 40-9 shows the patterns of the dual sense antenna.

For the C-141, the electrical center of the aircraft (pattern tilt angle of zero) was initially located through radiation-pattern measurements by using a small model configuration; a tilt-angle change was noted when the feed point for the cavity ADF antenna was moved. A zero tilt angle was obtained, and it was found that an approximate 10-in (254-mm) movement of the feed point aft produced a $+10^\circ$ tilt angle. Since a tilt angle of 0 to 10° is required, the measurements permitted selection of an optimum feed point ($+5^\circ$ tilt). Figure 40-9 shows the yaw, pitch, and roll radiation patterns obtained with the optimum antenna feed point.

The Rockwell/Collins DF-206A automatic direction finder (ADF) system is unique in that the DFA-206A antenna consists of a combined sense and loop antenna assembly. The sense antenna input is used as the reference input and is coupled to a low-pass filter. The low-pass filter removes the high-frequency interference. The two loop antennas are physically oriented 90° apart, and the resulting signal strengths are related to the direction of the antennas in relation to the transmitting station.

The antenna is a self-contained unit that may be located on either the top or the bottom of the aircraft. This unit is housed in a 16.5-in-long, 8.5-in-wide, and 1.75-in-thick teardrop designed to mount flush with the aircraft fuselage. It should be located as near the centerline

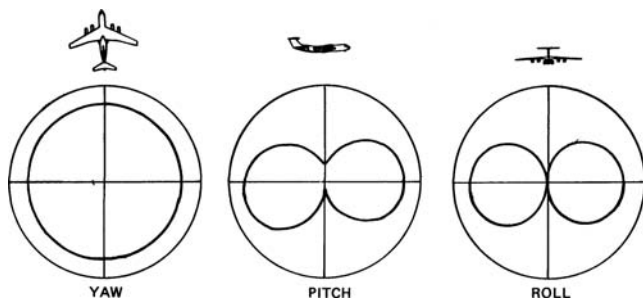


FIGURE 40-9 Voltage radiation patterns for the C-141 ADF sense antenna

of the aircraft as possible to reduce unsymmetric quadrantal error and near the wing root to obtain the best station-crossing indication. Unlike previous ADF systems, the length of the interconnecting cables between the antenna and receiver is neither limited nor critical. However, short cable runs are preferred to reduce the exposure to ground currents and interfering fields within the airframe.

40.3 HIGH-FREQUENCY COMMUNICATIONS ANTENNAS

Aircraft antennas for use with communications systems in the 2- to 30-MHz range are required to yield radiation patterns which provide useful gain in directions significant to communications. Also, impedance and efficiency characteristics suitable for acceptable power-transfer efficiencies between the airborne equipment and the radiated field are needed. A receiving antenna that meets these requirements will deliver to the input of a matched receiver atmospheric noise power under noise field conditions, which is many times greater than the input-circuit noise in the communications receiver. When this is the case, no improvement in signal-to-noise ratio can be achieved by further refinement of the antenna design. Since the transmitting mode of operation poses the more stringent requirements, the remainder of the discussion of HF antennas is confined to the transmitting case. Sky-wave propagation is always an important factor at these frequencies, and because of the rotation of polarization which is characteristic of reflection from the ionosphere, polarization characteristics are usually unimportant. Thus, the effective antenna gain can be considered in terms of the total power density. At frequencies below 6 MHz, ionosphere (and ground) reflections make almost all the radiated power useful for communication at least some of the time, so that differences between radiation patterns are relatively unimportant in comparing alternative aircraft antennas for communication applications in the 2- to 6-MHz range. In this range, impedance matching and efficiency considerations dominate. For frequencies above 6 MHz, pattern comparisons are frequently made in terms of the average power gain in an angular sector bounded by cones 30° above and below the horizon.

In the 2- to 30-MHz range, most aircraft have major dimensions of the order of a wavelength, and currents flowing on the skin usually dominate the impedance and pattern behavior. Since the airframe is a good radiator in this range, HF antenna design is aimed at maximizing the electromagnetic coupling to the airframe. The airframe currents exhibit strong resonance phenomena that are important to the impedance of antennas which are coupled tightly to the airframe.

Wire Antennas

Wire antennas, supported between the vertical fin and an insulated mast or trailed out into the airstream from an insulated reel, are reasonably effective HF antennas on lower-speed aircraft. Aerodynamic considerations limit the angle between a fixed wire and the airstream to about 15° . This means that fixed-wire antennas yield impedance characteristics similar to moderately lossy transmission lines, with resonances and antiresonances at frequencies at which the wire length is close to an integral multiple of $\lambda/4$. Figure 40-10 shows the input impedance of a 56-ft (17-m) fixed-wire antenna on a C-130 Hercules aircraft. Lumped reactances connected between the wire and the fin produce an effect exactly analogous to reactance-terminated lossy transmission lines. The average directive gain of these antennas in the sector $\pm 30^\circ$ relative to the horizon (for example, the fraction of the total radiated energy that goes into the sector, bounded by the cones 30° above and below the horizontal plane)

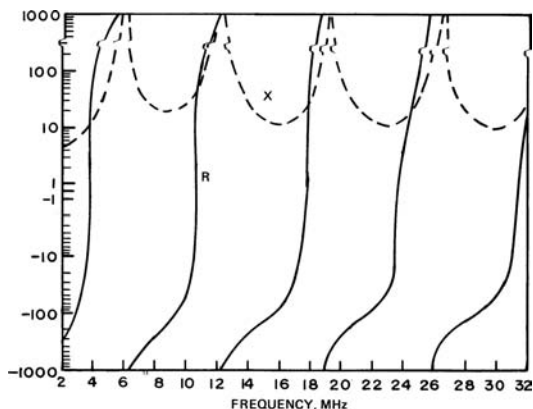


FIGURE 40-10 Impedance of the C-130E right-hand long-wire antenna

remains near 60 percent from 6 to 24 MHz. The efficiency of wire antennas is not high because of resistance loss in the wire itself and dielectric loss in the supporting insulators and masts. The resistance of commonly used wires is of the order of $0.05 \Omega/\text{ft}$ at 4 MHz. The RF corona breakdown threshold of fixed-wire antennas is a function of the wire diameter and the design of the supporting fittings. To minimize precipitation static, a wire coated with a relatively large-diameter sheath of polyethylene is frequently used, with special fittings designed to maximize the corona threshold. Even with such precautions, voltage breakdown poses a serious problem with fixed-wire antennas at high altitudes. Measurements indicate that standard antistatic strain insulators have an RF corona threshold of about 11 kV peak at an altitude of 50,000 ft (15,240 m)¹⁰ and a frequency of 2 MHz. Such an insulator, placed between the vertical stabilizer and the aft end of the open-circuited antenna, would go into corona at this altitude if the antenna were energized with a fully modulated AM carrier of about 150 W at 2 MHz.

Probe Antennas

A variation of the isolated-cap antenna is the probe or bullet-probe antenna. The radiation-pattern and impedance-resonance effects of the empennage, fuselage, and wings also affect probe antennas. The marked advantage of the probe antenna over the tail-cap antenna lies in the relative ease of incorporation into the structure of the aircraft empennage. The C-141 bullet-probe antenna shown in Figure 40-11¹¹ makes use of a portion of the structural intersection of the horizontal and vertical stabilizers in the T-tail configuration. The antenna is composed of the forward portion of the aerodynamic bullet with a probe extension. The plastic isolation gap in the bullet extends forward 11 in (279 mm), and the overall length of the antenna including the isolation band is 112 in (2845 mm). The maximum structural gap width was 13 in (330 mm), and the minimum gap width to withstand RF potentials was 3 to 4 in (76 to 102 mm). To obtain maximum airframe coupling, a maximum resistive component is required at the lowest expected resonant frequency. This resonance occurred at 2.4 MHz, which corresponds to a half-wavelength resonant condition caused by the wing-tip-to-horizontal-stabilizer-tip distance. Another parameter considered was the capacitive reactance of the antenna at 2.0 MHz. This reactance had to remain below 1000Ω if existing couplers were to be used. This resulted in

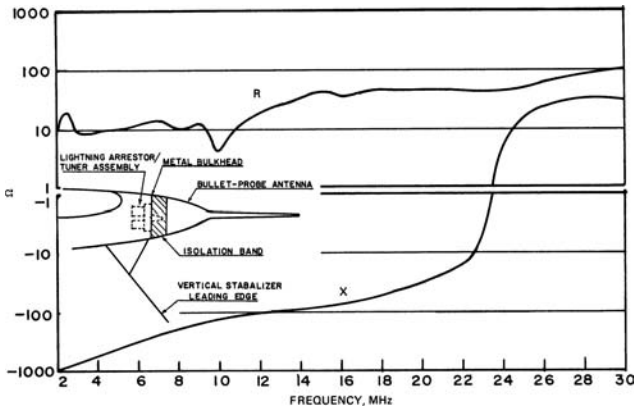


FIGURE 40-11 Impedance curves for the C-141 aircraft forward bullet-probe antenna¹¹

a gap width of 11 in for satisfactory performance. Figure 40-14 shows the impedance curves for the C-141 forward bullet-probe antenna.

A solid bulkhead was installed just aft of the isolation band for two reasons: (1) the metallic bulkhead acts as an electrostatic shield between the antenna and other metallic objects, and (2) the antenna tuner will mount on the bulkhead. Connection to the antenna is made through a lightning arrester which protrudes through the bulkhead into the gap area. The lightning arrester has a birdcage terminal which plugs into a flared tube, which is tied electrically to the probe and metal portions of the forward bullet.

Shunt or Notch High-Frequency Antennas

Shunt- and notch-type HF antennas are becoming increasingly popular on modern high-speed aircraft because of drag reduction and higher reliability. The basic difference in the shunt and notch configurations is the ratio of length to width of the cutout (dielectric) portion of the aircraft. Figure 40-12 shows the basic configurations.¹²

The shunt or notch antenna is a kind of transmission line that uses the vertical stabilizer or wing to feed the remainder of the aircraft. It can be located at various points on the airframe, but since it functions as an inductive coupler, it should be located at a point of low impedance or high current. The shunt-fed loop current causes the airframe itself to act as the antenna. The particular portion of the airframe supporting the current path is a function of the frequency and physical lengths of the parts of the airframe involved. For example, the wing-tip-to-horizontal-tip half-wavelength resonance on the C-5

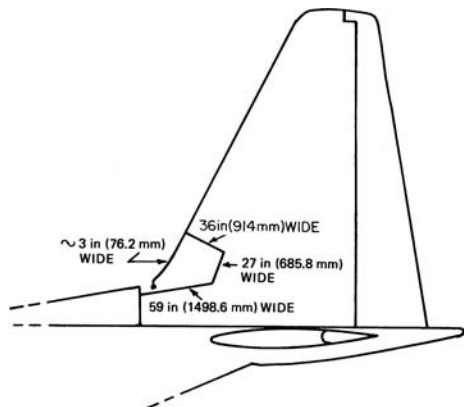


FIGURE 40-12 Basic configuration of the C-130 notch high-frequency antenna¹²

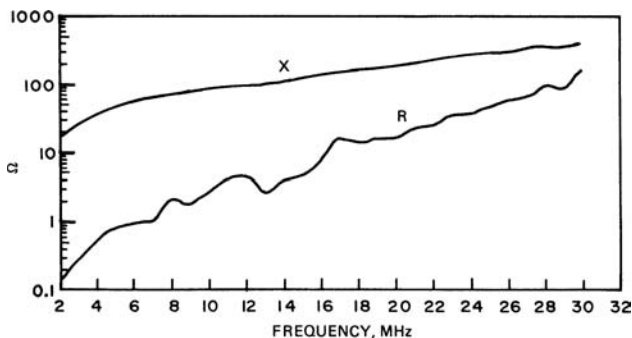


FIGURE 40-13 Impedance of the shunt-fed HF notch antenna on the C-130 aircraft¹²

aircraft occurs at approximately 2.5 MHz. The coupler feed point should always be located at the end of the shunt feed closest to the center of the aircraft.

The design of the antenna involves first finding a suitable location for the dielectric area and then arranging the dimensions of the shunt feed and dielectric area to provide an antenna that can be efficiently matched to $50\ \Omega$ with a relatively simple antenna coupler. Guidelines to observe are as follows: (1) The first choice for location on the airframe from a radiation-pattern and impedance standpoint is the root of the vertical stabilizer. An alternative is the wing root, but in this case the radiated vertically polarized energy is reduced. (2) The antenna coupler should be connected to the shunt feed near the stabilizer or wing root to provide maximum efficiency. (3) The dielectric material used to fill the opening should have low-loss characteristics. (4) The exact size and shape of the dielectric area is not a critical design parameter, but the dielectric area should be as deep as possible for maximum coupling to the airframe. Radiation efficiency is generally proportional to the dielectric area.

To avoid the need for a complicated antenna-tuning network, the antenna reactance at 2.0 MHz should be no less than $+j18\ \Omega$, and ideally parallel resonance should occur between 20 and 30 MHz. Series resonance should be completely avoided, and antenna reactance should be no lower than $-j100\ \Omega$ at 30 MHz. To ensure good coupling to the airframe, the parallel component of the antenna impedance should be less than $20,000\ \Omega$.

Figure 40-13 shows the shunt-fed notch antenna for the C-130 aircraft. A portion of the dorsal and a portion of the leading-edge fairing are combined to yield a notch in the root of the vertical stabilizer. The total distance around the notch is approximately 160 in (4064 mm), or slightly smaller than the ideal periphery of approximately 200 in (5080 mm). This was necessary because of interference with the leading-edge de-icing system. The impedance curve for this antenna is shown in Figure 40-14.

A unique application of the shunt-fed HF antenna is found on the Lockheed Jet Star aircraft.¹³ Figure 40-14 is a sketch of the antenna. The antenna is fed from a coaxial feed on the upper surface of the wing, where a 32-in (813-mm) piece of phosphor-bronze stranded cable connects the center conductor to the leading lower edge of an aft-mounted engine nacelle. Excellent omnidirectional patterns are obtained, and the impedance is tuned by using an off-the-shelf coupler inside the pressurized area of the fuselage.

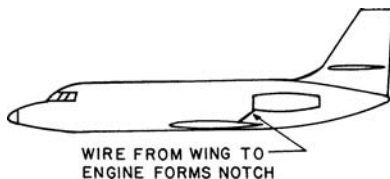


FIGURE 40-14 Shunt-fed HF notch-antenna configuration on the Lockheed JetStar¹³

40.4 UNIDIRECTIONAL VERY-HIGH-FREQUENCY ANTENNAS

The marker-beacon and glide-slope equipments require relatively narrowband antennas with simple patterns directed down or forward from the aircraft. This combination of circumstances makes the design of these antennas relatively simple. Both flush-mounted and external-mounted designs are available in several forms.

Marker-Beacon Antennas

The marker-beacon receiver operates on a fixed frequency of 75 MHz and requires a downward-looking pattern polarized parallel to the axis of the fuselage. A standard external installation employs a balanced-wire dipole supported by masts. The masts may be either insulated or conducting. Low-drag and flush-mounted designs are sketched in Figure 40-15. The low-drag design is a simple vertical loop oriented in the longitudinal plane of the aircraft. The feed line is inductively coupled to the loop, which is resonated by a series capacitor, and the antenna elements are contained in a streamlined plastic housing. The flush design is electrically similar, but in this case the structure takes the form of a conductor set along the longitudinal axis of the open face of a cavity. To achieve the desired impedance level, the antenna conductor is series-resonated by a capacitor and the feed point tapped partway along the antenna element.

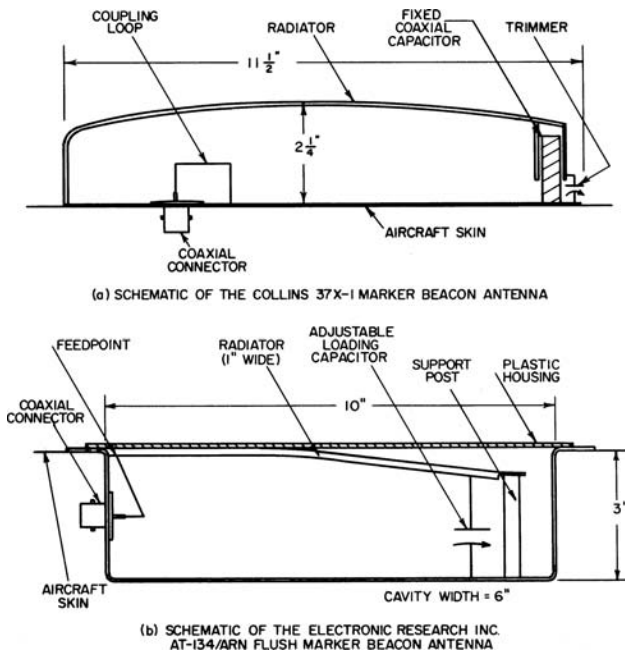


FIGURE 40-15 Typical marker-beacon antennas

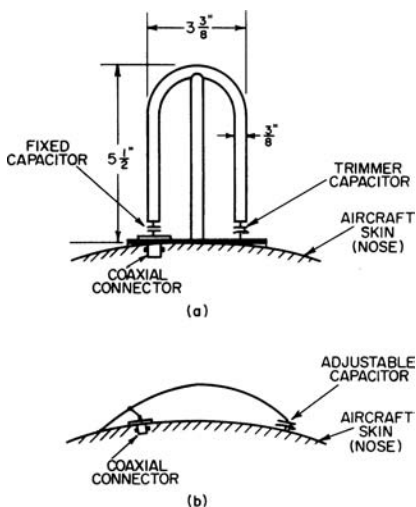


FIGURE 40-16 Protruding glide-slope antennas

Glide-Slope Antennas

The glide-slope receiver covers the frequency range from 329 to 335 MHz and requires antenna coverage only in an angular sector 60° on either side of the nose and 20° above and below the horizon. This requirement can be met by horizontal loops or by vertical slots. Because of the narrow bandwidth, the antenna element need not be electrically large. Two variations on the loop arrangement are sketched in Figure 40-16. Configuration *a* is a simple series-resonant half loop which can be externally mounted on the nose of the aircraft or within the nose radome. A variation of this antenna has two connectors for dual glide-slope systems. Configuration *b*, which is similar to the cavity-marker-beacon antenna of Figure 40-16, is suitable for either external or flush mounting.

40.5 OMNIDIRECTIONAL VERY-HIGH-FREQUENCY AND ULTRAHIGH-FREQUENCY ANTENNAS

In the frequency range in which the airframe is electrically large, the achievement of omnidirectional patterns, such as those used for short-range communications, is complicated by airframe effects. Since very-high-frequency (VHF) and ultrahigh-frequency (UHF) antennas of resonant size are structurally small, the required impedance characteristics can be achieved with fixed matching networks. Shadowing and reflection by the airframe result in major distortions of the primary pattern of the radiating element.

The tip of the vertical fin is a preferred location for omnidirectional antennas in the VHF and UHF range because antennas located there have a relatively unobstructed view of the surrounding sphere. At these frequencies the mean chord of the fin is approximately 10 wavelengths and the principal effect of the airframe on the radiation patterns is a sharply defined optical shadow region. At the lower end of this frequency range and on smaller aircraft, the effect of the airframe on the patterns is more complicated. Deep nulls in the forward quadrants can occur because of the destructive interference between direct

radiation and radiation reflected from the fuselage and wings. The latter contribution is more important for small aircraft since the ground plane formed by the surface of the fin tip is now sufficiently small enough to permit strong spillover of the primary pattern, which also can create lobing in the transverse-plane pattern. In this frequency range, the null structure is strongly influenced by the position of the radiating element along the chord of the fin, and careful location of the antenna along the chord may result in improvements in the forward-horizon signal strength.

UHF/VHF Antennas for Vertical Polarization

The monopole and its variants are the most commonly used vertically polarized VHF and UHF aircraft antennas. The antenna typically has a tapered airfoil cross section to minimize drag. Simple shunt-stub matching networks are used to obtain a voltage standing-wave ratio (VSWR) below 2:1 from 116 to 156 MHz.

The basic configuration of an all-metal VHF blade antenna is shown in Figure 40-17. This antenna employs a shunt-fed slot to excite the blade. The advantage of this configuration is its all-metal construction, which provides lightning protection and eliminates p static. Figure 40-18 shows a low-aerodynamic-drag sleeve monopole for the 225- to 400-MHz UHF communications system. The cross section is diamond-shaped, with a thickness ratio of 5:1. A feature of this antenna¹⁴ is the introduction into the impedance-compensating network of a shunt stub such that its inner conductor serves as a tension member to draw the two halves of the antenna together. In addition to providing mechanical strength, this inner conductor forms a direct-current path from the upper portion of the antenna to the aircraft skin, thereby providing lightning-strike protection.

Figure 40-19 is a broadband monopole/blade antenna matched through a pair of shunt blocks and a series feed with a capacitive section of semirigid coax. The shunt blocks are similar to L/R shunts and are ferrite-based for a gain of -24.5 dBi. There is an abrupt drop-off of antenna performance below 30 MHz.

A number of monopole designs have been developed for installation on a fin in which the top portion of the metal structure has been removed and replaced by a suitable dielectric housing. Various other forms of vertically polarized UHF and VHF radiators have been designed for tail-cap installation. The pickax antenna, consisting of a heavily top-loaded vertical element, has been designed to provide a VSWR of less than 3:1 from 110 to 115 MHz with an overall height of $15\frac{1}{2}$ in (393.7 mm) and a length of 14 in (355.6 mm).

UHF/VHF Antennas for Horizontal Polarization

There are three basic antenna elements which yield omnidirectional horizontally polarized patterns: the loop, the turnstile, and the longitudinal slot in a vertical cylinder of small diameter. All three are used on aircraft, and all suffer from a basic defect. Because they must be mounted near a horizontal conducting surface of rather large extent (i.e., the top or bottom surface of the aircraft), their gain at angles near the horizon is low. The greater the spacing from the conducting surface, the higher the horizontal gain. For this reason, locations at or near the top of the vertical fin are popular for horizontally polarized applications, particularly for the VHF navigation system (VOR) which covers the 108- to 122-MHz range.

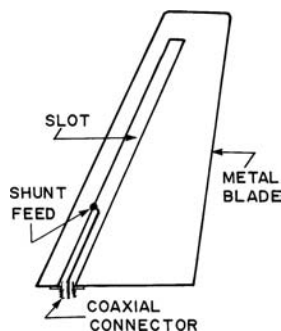


FIGURE 40-17 Basic configuration of an all-metal VHF blade antenna

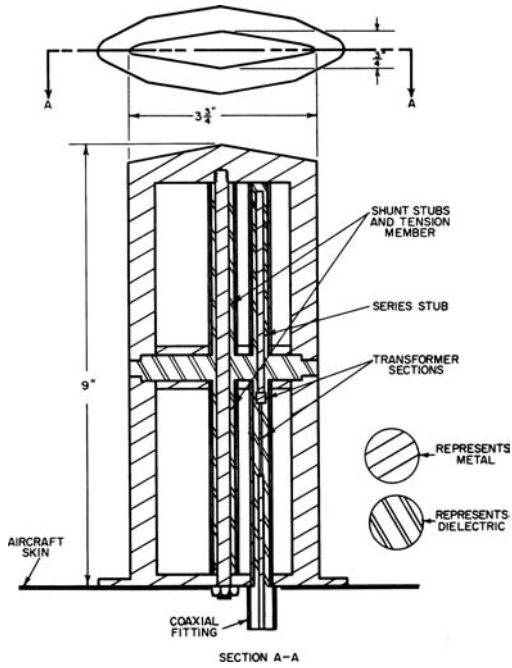


FIGURE 40-18 The AT-256A antenna

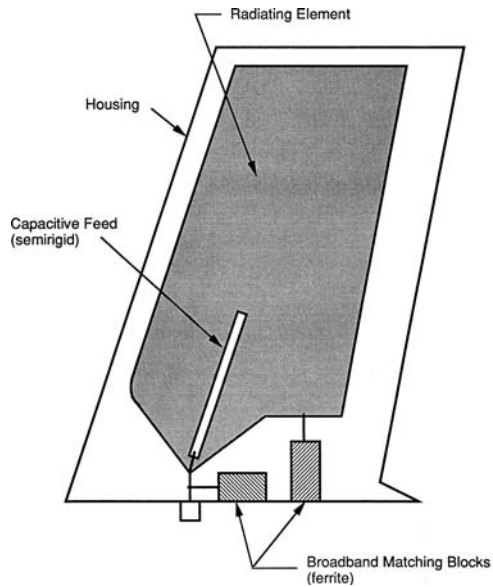


FIGURE 40-19 Typical broadband (30–400 MHz) blade antenna, model DMC144 (*Courtesy of Dorne & Margolin, Inc.*)

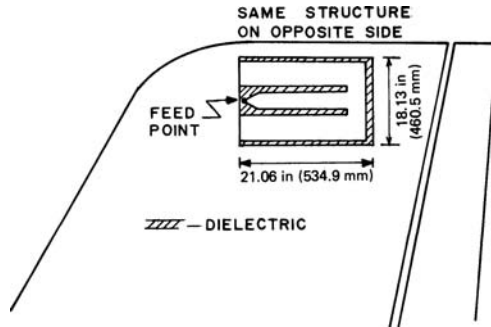


FIGURE 40-20 *E*-slot VOR antenna for the L-1011 aircraft (after L. Schuchman et al¹⁸ © IEEE 1989)

Figure 40-20 shows an *E*-fed cavity VOR antenna designed into the empennage tip of the L-1011 aircraft.¹⁵ The antenna system consists of a flush-mounted dual *E*-slot antenna and a stripline feed network mounted on the forward bulkhead of the cavity between the two antenna halves. The antenna halves are mirror-image assemblies consisting of 0.020-in- (0.508-mm-) thick aluminum elements bonded to the inside surface of honeycomb-fiberglass windows. The feed network consists of a ring hybrid-power-divider device with four RF ports fabricated in stripline. The two input ports for connecting to the antenna elements are out-of-phase ports. The two output ports connect to separate VOR preamplifiers. The operating band of the antenna is 108 to 118 MHz with an input VSWR of less than 2:1 referred to 50 Ω . Radiation coverage is essentially omnidirectional in the horizontal plane. The principal polarization is horizontal with a cross-polarization component of more than 18 dB below horizontal polarization. The antenna gain is 3.6 dB below isotropic at band center. Figure 40-21 shows the principal-plane voltage radiation patterns.

Figure 40-22 is a single half-loop antenna which is commonly installed on each side of the vertical stabilizer. The installed two-half-loop antenna system is basically a forced-fed balanced loop. The electrical equivalent is a single loop with a single feed with no intervening ground plane. The feed points are 180° out of phase to create a circuitous current flow with cancellation in the plane of the vertical stabilizer. Matching is achieved through one inductive and one capacitive shunt, and the feed gap is joined through a capacitive feed.

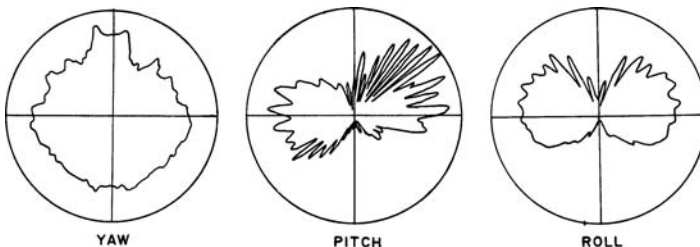


FIGURE 40-21 Voltage radiation patterns for the *E*-slot VOR antenna on the L-1011 aircraft (after L. Schuchman et al¹⁸ © IEEE 1989)

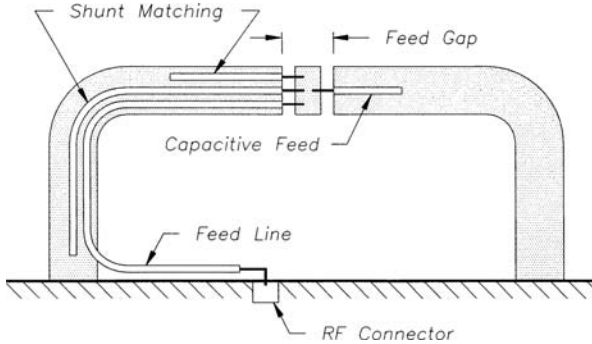
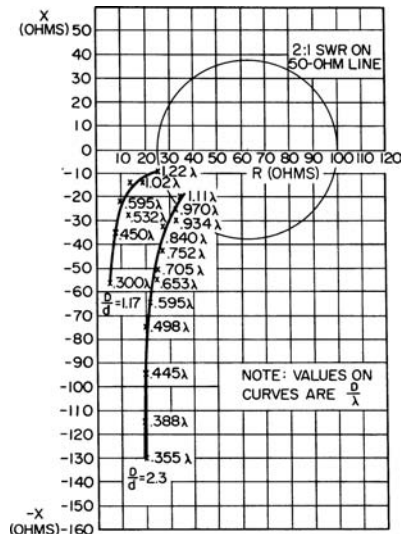


FIGURE 40-22 Single half-loop antenna, model DMN4-33 (Courtesy of Dorne & Margolin, Inc.)

40.6 OMNIDIRECTIONAL L-BAND FREQUENCY ANTENNAS

The monopole and its variants are also commonly used for vertically polarized L-Band aircraft antennas. The antenna designs used for VHF and UHF frequencies can be scaled down to resonate at higher frequency ranges. Flush-mounted antenna designs are also possible since antenna dimensions are small enough to fit between aircraft structural elements.

The basic flush-mounted vertically polarized element for fuselage mounting is the annular slot,¹⁶ which can be visualized as the open end of a large-diameter, low-characteristic-impedance coaxial line. As seen from the impedance curve of Figure 40-23, such a structure becomes an effective radiator only when the circumference of the slot approaches a wavelength. The radiation patterns have their maximum gain in the plane of the slot for very small



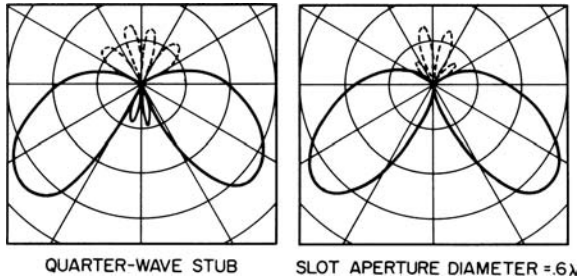


FIGURE 40-24 Radiation patterns of an annular-slot antenna and a quarter-wave stub on a $2\frac{1}{2}$ -wavelength-diameter circular ground plane

slot diameters only, and they yield a horizon gain of zero for a slot diameter of 1.22.¹⁷ This pattern variation is illustrated in Figure 40-24. For these reasons and to minimize the structural difficulty of installing the antenna in an aircraft, the smallest possible diameter yielding the required bandwidth is desirable.

Again, the design for a UHF annular-slot antenna can be scaled down for use at L-Band. Consider a UHF broadband design for the 225- to 400-MHz band. The minimum practical diameter, considering construction tolerances and the effect of the airframe on the impedance, is about 24 in (610 mm). This is a rather large hole to put in an aircraft fuselage and is why this type of antenna is typically used at L-Band frequencies and higher. A VSWR under 2:1 across the 225- to 400-MHz band can be obtained with this diameter and a cavity depth of 4.5 in (114.3 mm). Figure 40-25 shows a design by A. Dörne together with its approximate equivalent circuit.

In the equivalent circuit the net aperture impedance of the driven annular slot and the inner parasitic annular slot is shown as a series resistance-capacitance (RC) circuit. The annular region 1, which is coupled to the radiating aperture through the mutual impedance between the two slots, and the annular region 2, which is part of the feed system, are so positioned and proportioned that they store primarily magnetic energy. The inductances associated with the energy storage in these regions are designated as L_1 and L_2 in the equivalent circuit. The parallel-tuned circuit in the equivalent circuit is formed by the shunt capacitance between vane 3 and horizontal disk 4, together with the shunt inductance provided by four conducting posts (5) equally spaced about the periphery of 4, which also serve to support 4 above 3. From this element inward to the coaxial line, the base plate is cambered upward to form a conical transmission-line region of low characteristic impedance. A short section of low-impedance line is added external to the cavity to complete the required impedance transformation.

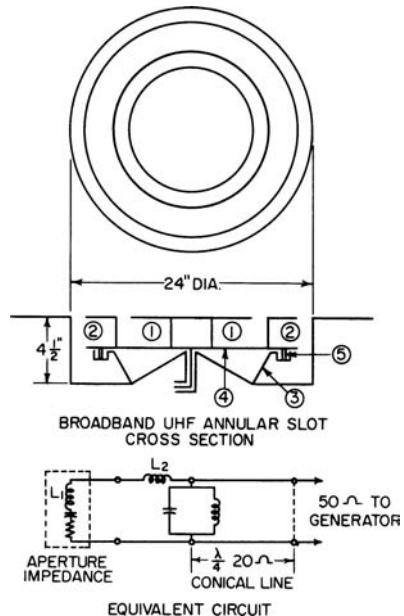


FIGURE 40-25 Annular-slot antenna for the 225- to 400-MHz band and its equivalent circuit

40.7 GPS ANTENNAS

The deployment of the Global Positioning System (GPS) “promises unprecedented world-wide, all-weather navigation capabilities when fully implemented in the 1990s.”¹⁸ The GPS system certainly provides general navigation functions and may be extended for Category 1 through 3 landings by the use of Differential GPS (DGPS) techniques.^{18,19} This technology is currently being evaluated, as seen in “Comments on Alternatives to the Microwave Landing System (MLS).”¹⁹ The GPS satellite transmits information on two frequencies: L_1 at 1575.42 MHz and L_2 at 1227.6 MHz. The L_1 frequency carries a code known as the *C/A code* and a *P-code*. The L_2 frequency carries only the P-code, which is controlled by the Department of Defense and is encrypted for use of classified military users.²⁰ Thus GPS antennas are required for both civil and military applications.

In order to obtain maximum position accuracy, a field of view which encompasses four satellites is required. In general, this dictates a near-hemispherical antenna pattern with a coverage extending from near the horizon to the zenith. However, the ideal antenna pattern would maximize horizon gain at the expense of zenith gain due to geometry considerations. That is, the satellites at the horizon are at the maximum range. The GPS satellite transmits right-hand circular polarization. Therefore, a linearly polarized antenna can be employed with a 3-dB gain-loss penalty. Although maximum gain is desirable, low antenna gain is compensated for by the processing gain inherent in the GPS code. Typical GPS antenna gains range from -7.5 to -3 dBci at the horizon to 0 to 4 dBci at zenith, depending on the antenna utilized. For civilian applications, the typical antennas consist of crossed slots, spirals, annular slots, microstrip quadrafilers, helices, and microstrip patches. The patch antennas typically consist of stacked patches which are tuned to the two frequencies. These antennas are categorized as fixed reception pattern antenna (FRPA). For military applications, which may utilize adaptive null steering to counter interfering jamming signals, a controlled reception pattern antenna (CRPA) may be used. These antennas are typically microstrip patch arrays which are capable of adaptive beam forming.

40.8 ALTIMETER ANTENNAS

The radio altimeter, which operates at 4300 MHz, requires independent downward-looking antennas for transmission and reception. Proper operation requires a high degree of isolation between the transmitting and receiving elements. Horn antennas were originally used, but microstrip patch antennas are most commonly used today due to their smaller size, weight and structural modification requirements.

40.9 ANTENNA PERFORMANCE AND AIRCRAFT INTERACTION

Several techniques are used in characterizing the impedance and radiation pattern characteristics of an antenna installed on an aircraft. Scale model testing has historically been the most common method used for radiation pattern measurements. Full-scale measurements are expensive, require enormous facilities, and are therefore usually employed only in large budget government programs. Computer simulations have matured with the growth in processing power and are widely used today. All of these techniques can be used to predict the aircraft's effects on antenna performance, including modulation of the radiated signal due to motion of the blades in the case of propeller-driven aircraft and helicopters.

Scale Model Pattern Measurements

Using scale models overcomes many of the difficulties associated with radiation pattern measurements. The distance to the far-field can be reduced so that a reasonably sized indoor or outdoor pattern range can be used. A properly designed range can manipulate the model of the aircraft with the antenna to get the desired pattern measurements and provide a controlled environment to help eliminate unwanted reflections from the ground and surrounding objects or allow their removal through signal processing techniques. When developing a scale model, if the aircraft and antenna are reduced in size by a given scaling factor, the frequency and conductivity must be increased (multiplied) by that same factor. The accuracy of the model in terms of detail and electrical parameters depends on the frequency of operation, location of the antenna relative to the detailed parts, and required accuracy of the pattern performance predictions.

Most external antennas are located on the top or bottom centerline of the fuselage in order to maintain symmetry of the radiation patterns. Pattern coverage in such locations is limited by the airframe shadows. Figure 40-26 shows the scale model patterns for a monopole on the bottom centerline of a one-fifth scale C-141 model. Since the aircraft is one-fifth scale, the frequency measured is five times greater and the monopole is designed to resonate at 1000 MHz. It is apparent that coverage is limited to the hemisphere below the aircraft. Antennas on the top of the fuselage yield patterns confined to the upper hemisphere, with a null aft owing to the shadow cast by the vertical stabilizer. In many applications, as for example in scheduled-airline operations, these pattern limitations are acceptable.

The deep lobing in the roll-plane pattern of Figure 40-26 is due to reflections from the strongly illuminated engine nacelles. In some locations, similar difficulties are encountered because of reflections from the wing flaps when they are extended. Shadows and lobing also occur due to extended landing gear and can be troublesome for bottom-mounted antennas.

Full-Scale Pattern Measurements

If a full-scale version of the aircraft and antenna can be accommodated by an antenna pattern range, detailed performance characteristics can be measured. For example, the VOR navigation system is particularly vulnerable to the modulation effects of a helicopter rotor. Angular-position information is contained in a 30-Hz modulation tone which corresponds closely to the third harmonic of the fundamental blade-passage frequency on a typical helicopter. In Figure 40-27, two VOR antenna installations on an H-19 helicopter are shown with the horizontal-plane radiation patterns.

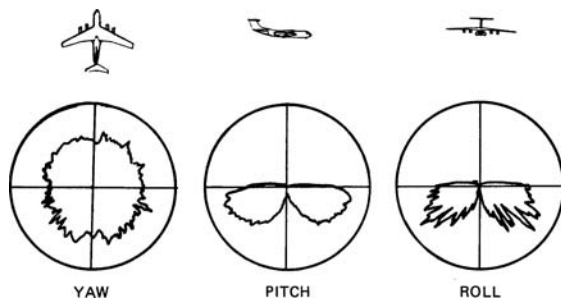


FIGURE 40-26 Voltage radiation patterns of a 1000-MHz monopole on the bottom of a C-141 aircraft

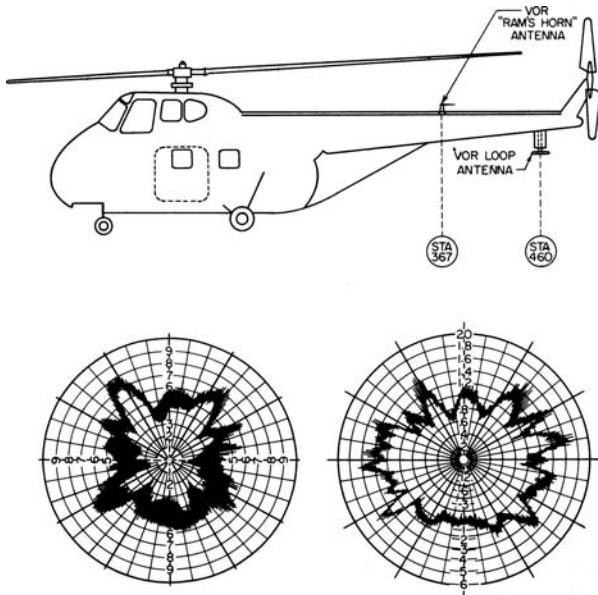


FIGURE 40-27 Two VOR-antenna locations on the H-19 helicopter and horizontal-plane radiation patterns

The fine structure on these patterns shows the peak-to-peak variation in signal amplitude due to passage of the rotor blades. The percentage of modulation of the signal received on the horizontal loop antenna is seen to be lower than that received on the ramshorn antenna. This is due partly to the shielding afforded the loop by the tail boom and partly to the fact that the loop antenna has inherently less response to scattered signals from the blades because its pattern has a null along its axis while ramshorn antenna receives signals from directly above it very effectively.

Method-of-Moments Analysis of Low-Frequency and High-Frequency Antenna Patterns

The method of moments (MOM)¹⁻³ is a computer analysis method which can be employed to determine the radiation patterns of antennas on aircraft and missiles. A discussion of this technique is beyond the scope of this handbook, but user-oriented computer programs are available.²¹⁻²³

The method of moments is based upon a solution of the electromagnetic-wave-scattering integral equations by enforcing boundary conditions at a number N of discrete points on the scattering surface. The resulting set of equations is then solved numerically to determine the scattered fields. This is typically accomplished by modeling the surface as a wire grid for which the self- and mutual impedances between the wires can be calculated. This technique, for a set of N wire elements that represent the aircraft as shown in Figure 40-28, results in an $N \times N$ impedance matrix. For a given antenna excitation, inversion of the $N \times N$ matrix yields the solution for the wire currents and hence the radiated fields of the antenna-aircraft system.

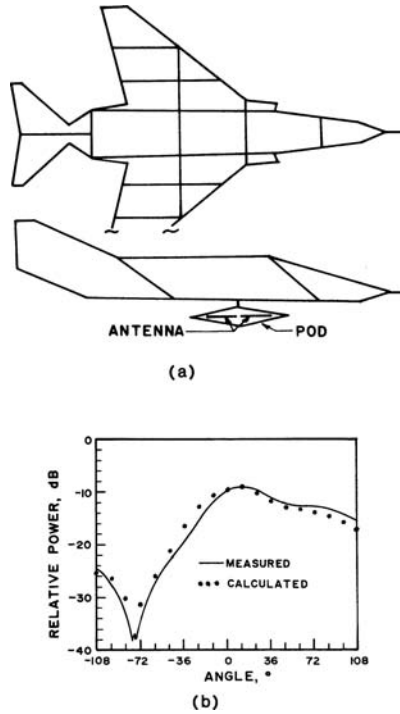


FIGURE 40-28 MOM wire-grid analysis of low-frequency aircraft antenna patterns: (a) MOM wire-grid model of an RF-4C aircraft; (b) MOM calculated roll-plane pattern for a pod-mounted dipole array on the RF-4C aircraft (after J. J. Wang and C. E. Ryan²⁴ © IEEE 1977)

Computer programs which implement this technique for wire scatterers^{1,21-24} and for surface patches^{25,26} have been written. Supercomputers can handle matrixes on the order of 10,000 by 10,000 elements, and continuing technology advances will permit larger matrixes to be solved. Also, several MOM codes for personal computers (PCs) are available.^{22,23} These codes will become increasingly useful as the computing capability of PCs increases. The number of wire or patch elements which must be employed to ensure a convergent numerical solution varies as a function of the particular code implementation. Typically, a wire element less than $\lambda/8$ in length must be used so that $1/\lambda^2$ surface patch requires a square wire-grid model containing 144 wires. Thus it is seen that as the surface area increases, the required matrix increases roughly as the square of the area which, in practice, restricts the MOM to electrically small- to moderate-sized geometries. Codes based on body-of-revolution approximations are also available²⁵ and are particularly applicable to missiles.

A wire-grid model which was used for the pattern analysis of a pod-mounted high-frequency (HF) folded-dipole array antenna on an RF-4C aircraft is shown in Figure 40-28. An MOM code employing piecewise, sinusoidal basis functions^{1,24} was used to calculate the roll-plane patterns for the array. These results are compared with measured data in Figure 40-28.

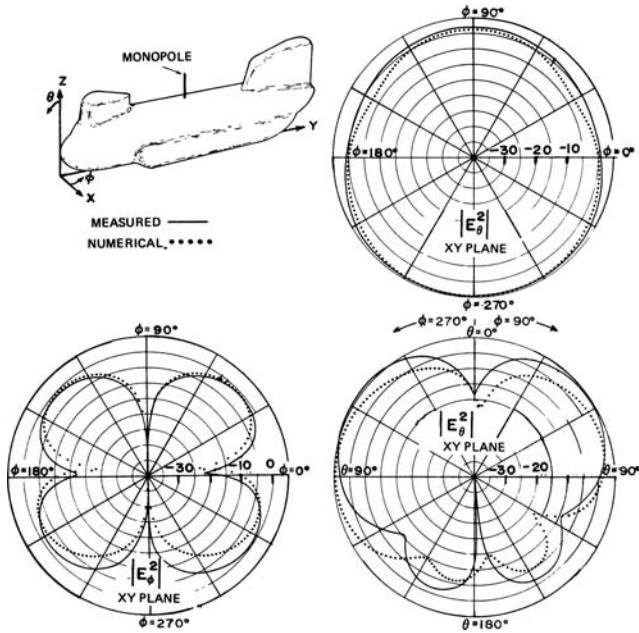


FIGURE 40-29 MOM surface-patch-calculated patterns for the CH-47 helicopter (after D. L. Knepp and J. Goldhirsh²⁷ © IEEE 1972)

A surface-patch method has also been employed to analyze the patterns of antennas on helicopters.²⁷ In this analysis, the helicopter surface was subdivided into a collection of curvilinear cells. A result of the analysis, which employed 94 surface-patch cells, is shown in Figure 40-29 for the pitch and yaw planes. Generally, good results were obtained for this complex-shaped airframe. Also, we note that, by representing the helicopter rotor blades as wires, the rotor effects on the patterns can be computed.

For some LF applications, a simplified stick model of an aircraft may be sufficient.²⁸ A model of an aircraft is shown in Figure 40-30. The fuselage, wings, and empennage can be represented by either single or multiple thin wires if their maximum electrical width

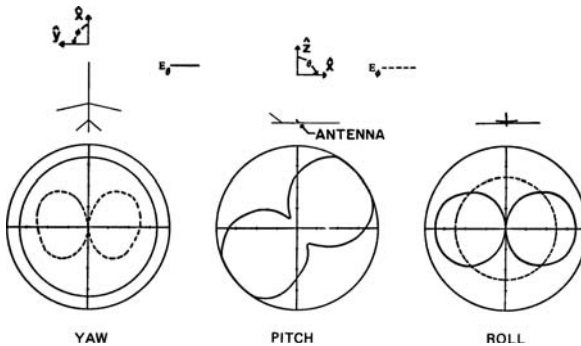


FIGURE 40-30 MOM calculated voltage patterns for a blade antenna (tilted monopole) on a stick model aircraft

is approximately less than $\lambda/10$. Figure 40-12 shows a calculated result for a monopole mounted on a stick model, and the effects of the fuselage and wings can be seen by comparison with the unperturbed-monopole pattern.²⁹

GTD Aircraft Antenna Analysis

The MOM technique is limited to surfaces which are relatively small in terms of square wavelengths. However, as the frequency increases and the surface becomes large in terms of square wavelengths, the propagation of electromagnetic (EM) energy can be analyzed by using the techniques of geometrical optics. For example, if an incident field is reflected from a curved surface, ray optics and the Fresnel reflection coefficients yield the reflected field.

The geometrical theory of diffraction (GTD) uses the ray-optics representation of EM propagation and incorporates both *diffracted rays* and *surface rays* to account for the effects of edge and surface discontinuities and surface-wave propagation. In the case of edge diffraction, the diffracted rays lie on a cone whose half angle is equal to the half angle of the edge tangent and incidence vector. For a general curved surface, the incident field is launched as a surface wave at the tangent point, propagates along a geodesic path on the surface, and is diffracted at a tangent point toward the direction of the receiver. Each of these diffraction, launching, and surface propagation processes is described by appropriate complex diffraction and attenuation coefficients which are discussed in the literature.³⁰ In practice, these coefficients are implemented as computer algorithms which can be used as building blocks for a specific analysis.

In the GTD analysis, the received far- or near-zone field is composed of rays which are directly incident from the antenna, surface rays, reflected rays, and edge-diffracted rays. Also *higher-order effects* such as multiple reflections and diffractions can occur. Several computer models have been implemented to perform the required differential-geometry and diffraction calculations.³¹⁻³³ It is anticipated that these methods will become more user-oriented and powerful with the advent of increasingly sophisticated computer-graphics systems.

Figure 40-21 shows a model of a missile composed of a conically capped circular cylinder with a circumferential-slot antenna.³⁴ In this case, the radiated field is due to direct radiation from the slot, surface rays which encircle the cylinder, and diffracted rays from the cone-cylinder and cylinder-end junctions and from the cone tip. The computed results shown in Figure 40-31 were obtained by employing the solution for a slot on an infinite circular cylinder to represent the direct- and surface-ray fields and by computing the diffracted fields from the junctions and tip as corrections to the infinite-cylinder result. The principal-plane pattern shown is in agreement with measured data, and comparable results have been obtained for the off-principal planes.³⁵

An analysis due to Burnside et al.³¹ approximates the fuselage as two joined spheroids, and the wings and empennage are modeled by arbitrarily shaped flat plates. In addition, the engines can be modeled by finite circular cylinders. Figure 40-32 shows a result obtained from this analysis of a $\lambda/4$ monopole on a KC-135 aircraft for the roll plane. J. R. Mantz and R. F. Harrington²⁵ present a complete 4π -sr plot for a monopole on a Boeing 737 aircraft. The results, which compare well with measured data, illustrate the accuracy of the GTD analysis for VHF and UHF antennas.

It should be noted that a complete GTD model may not always be required for an engineering assessment of antenna performance. For example, if wing reflection and blockage effects are dominant, one can use a simpler model consisting of two submodels: a submodel for an antenna on a finite (or perhaps an infinite) cylindrical fuselage and a flat-plate submodel for the wings.³³ In this case, the antenna-on-cylinder pattern results are used as the antenna illumination for calculation of the wing blockage, reflection, and diffraction.

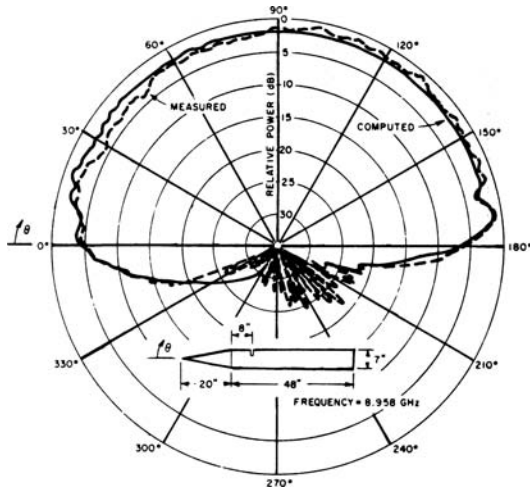


FIGURE 40-31 Pattern of a circumferential slot on a conically capped circular cylinder (after D. L. Knepp and J. Goldhirsh²⁷ © IEEE 1972)

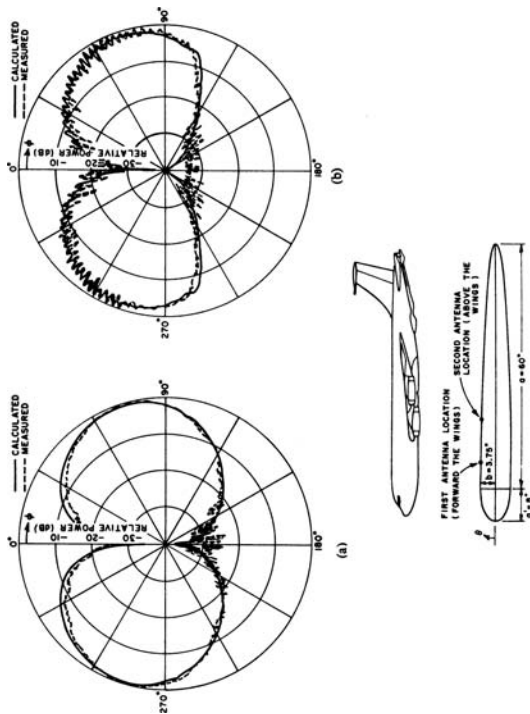


FIGURE 40-32 (a) Roll-plane pattern (E_θ) for a 1:25-scale model of a KC-135 aircraft with a $\lambda/4$ monopole on the fuselage forward of the wings at a frequency of 34.92 GHz (model frequency); (b) Roll-plane pattern (E_θ) for a $\lambda/4$ monopole over the wings (after W. D. Burnside et al³¹ © IEEE 1975)

The computer analysis of aircraft interference effects on scanning array antenna patterns is a subject of much interest to those working in the field of large airborne antenna arrays. Experience indicates that such analyses employing the geometrical theory of diffraction (GTD) are feasible and cost-effective. The emphasis in the technology is the use of large-aperture, low-frequency (UHF to S-band) arrays installed on large airframes such as the C-5A or B747 aircraft. Techniques usually considered are scale-model measurements, full-scale measurements, and computer simulation of the array antennas on airframes. Computer-simulation techniques have a considerable advantage from a cost standpoint over the other two options. The computer-simulation technique is a very useful tool for determining the proper array location on the airframe and the frequency sensitivity to airframe effect on radiation patterns. Wing, engine nacelles, and empennage structure in the near field of large arrays can cause devastating effects on the sidelobe structure of a well-designed array antenna. Once the airframe and the array are simulated on the computer, it is a simple matter to relocate the array on the airframe and compute the resulting pattern. While the pattern computation is very central process unit (CPU) time-intensive, it is still more cost-effective to move the array on the computer than to move the array on the full-scale aircraft or scale model.

A collection of plates and cylinders is used to model the airframe, including the empennage, wing, engine nacelles, and engine pylons. Each element in the array antenna is assumed to be a separate source controllable in amplitude and phase, allowing the use of aperture-distribution tapering and simulated electronic scanning of the beam to a desired angle.

This technique allows the computation of three parameters to help with the evaluation of antenna performance on the airframe. They are the two-way integrated sidelobe-to-mainlobe (SL/ML) ratio in decibels, the rms sidelobe level in decibels, and the 3-dB beamwidth in degrees. To calculate the SL/ML ratio, the sidelobes and mainlobe are sampled at 1° increments and integrated. The integrated sidelobe power (two-way) is divided by the integrated mainlobe power (two-way). One of the great strengths of this computer-analysis technique is the user's ability to disassemble the airframe, when desired, for identifying the primary contributor to a particular sidelobe or cross-polarized component of the radiation patterns.

The computer code used to compute phased-array antenna radiation patterns is a modified version of the "Numerical Electromagnetic Code—Basic Scattering Code" (version 3.19) developed by the Ohio State University ElectroScience Laboratory.^{36,37} It uses the geometrical theory of diffraction (GTD) to determine the total field at a given observation point. Several basic geometric shapes can be used to model scattering objects. They are arbitrarily oriented flat plates, perfectly conducting finite elliptic cylinders, elliptic cone frustum sections, and finite composite ellipsoids. These objects should have a minimum dimension of one wavelength, and the accuracy increases as the object size in wavelengths increases. The required accuracy of the scattering model also depends on the direction of the mainbeam, while broad-beam antennas require greatest accuracy in the part of the model near the antenna.

The GTD code was used to analyze a rotodome array installed on the S3 aircraft, as shown in Figure 40-33. The 28-in by 228-in array is composed of 32 columns spaced 7.125 in by 4 rows spaced 7.0 in. A 50-dB Taylor ($n = 6$) aperture distribution was chosen for the horizontal plane and a 30-dB Taylor ($n = 4$) aperture distribution was chosen for the vertical plane. The free-space azimuth plane pattern, shown in Figure 40-34, has a 3-dB beamwidth of 4.75° in the horizontal plane at 900 MHz. When this array is installed in the rotodome above the S3 airframe, the resulting pattern for a 45° look angle from the nose of the aircraft is shown in Figure 40-35. The integrated two-way SL/ML ratio for the free-space radiation pattern is -91 dB. When the array is installed on the S3 airframe, the SL/ML ratio is -71.6 dB. This parameter is indicative of how severe the competing sidelobe clutter might be in a radar application. The acceptability of this much sidelobe degradation would be determined by the radar designer.

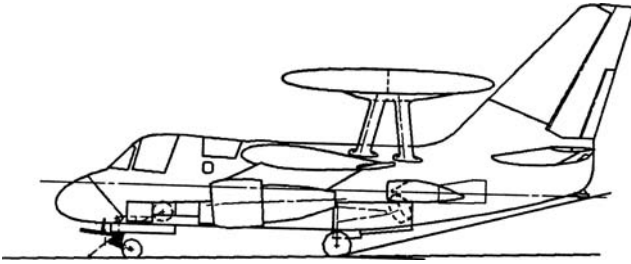


FIGURE 40-33 S3 aircraft showing rotodome installation (Courtesy of Lockheed Aeronautics Systems Company)

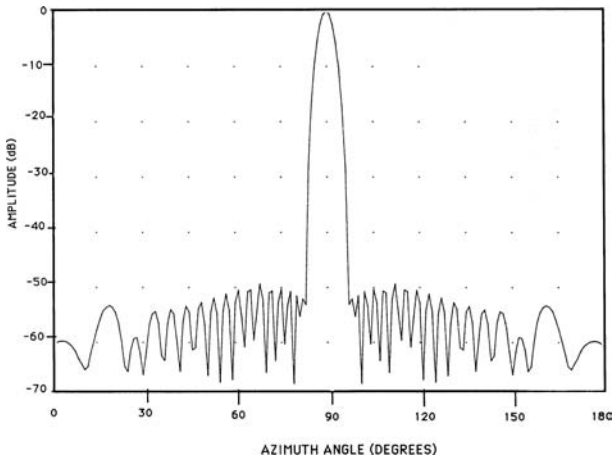


FIGURE 40-34 Free-space azimuth plane array pattern (Courtesy of Lockheed Aeronautics Systems Company)

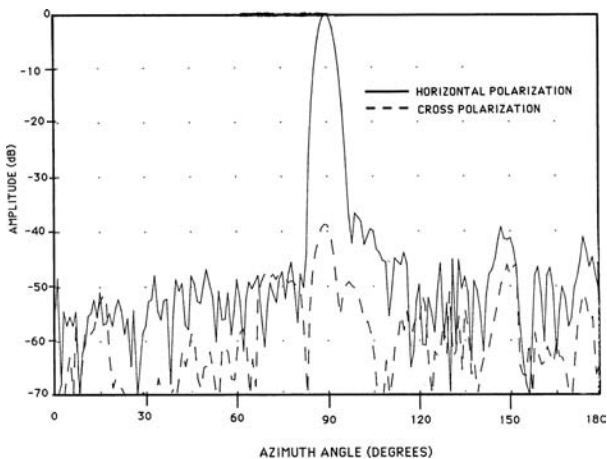


FIGURE 40-35 Azimuth plane pattern with the array installed above the S3 airframe with a 45° azimuth look angle (Courtesy of Lockheed Aeronautics Systems Company)

REFERENCES

1. J. H. Richmond, "A Wire-Grid Model for Scattering by Conducting Bodies," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (November 1976): 782–786.
2. R. F. Harrington, *Field Computation by Moment Methods* (New York: The Macmillan Company 1968).
3. J. J. Wang, *Generalized Moment Method in Electromagnetics* (New York: John Wiley, 1991).
4. *Instruction Book for Type NA-1 Aircraft Navigation System* (Baltimore, Maryland: Bendix Radio Div., Bendix Aviation Corp.).
5. L. R. Mullen, "The Marconi AD 7092 Series of ADF Receivers," *IRE Trans. Aeronaut. Navig. Electron.*, vol. AN-2, no. 4 (December 1955).
6. J. T. Bolljahn and R. F. Reese, "Electrically Small Antennas and the Low-Frequency Aircraft Antenna Problem," *IRE Trans. Antennas Propagat.*, vol. AP-1, no. 2 (October 1953): 46–54.
7. A. A. Hemphill, "A Magnetic Radio Compass Antenna Having Zero Drag," *IRE Trans. Aeronaut. Navig. Electron.*, vol. AN-2, no. 4 (December 1955).
8. H. H. Ward, 3d, "Analysis of the Over-Station Behavior of Aircraft Low-Frequency ADF Systems," *IRE Trans. Aeronaut. Navig. Electron.*, vol. AN-2, no. 4 (December 1955): 31–41.
9. P. M. Burdell, "C-141A ADF Sense Antenna Development," Eng. Rep. 5832, Lockheed-Georgia Company, May 28, 1963.
10. R. L. Tanner, "High Voltage Problems in Flush and External Aircraft HF Antennas," *IRE Trans. Aeronaut. Navig. Electron.*, vol. AN-1, no. 4 (December 1954): 16–19.
11. B. S. Zieg and W. P. Allen, "C-141A HF Antenna Development," Eng. Rep. ER 5101, Lockheed-Georgia Company, July 1963.
12. P. M. Burdell, "C-130 HF Notch Antenna Design and Development," Eng. Rep. LG 82 ER 0036, Lockheed-Georgia Company, March 1982.
13. B. S. Zieg, "JetStar HF Antenna," Tenth Ann. USAF Antenna R & D Symp., University of Illinois, Monticello, Illinois, October 3–7, 1960.
14. H. Jasik, U.S. Patent 2,700,112.
15. N. R. Ray, "L-1011 VOR Antenna System Design and Development," Eng. Rep. ER-10820, Lockheed-Georgia Company, August 13, 1970.
16. A. Dorne, U.S. Patent 2,644,090.
17. A. A. Pistolcors, "Theory of the Circular Diffraction Antenna," *IRE Proc.*, vol. 36, no. 1 (January 1948): 56.
18. L. Schuchman, B. D. Elrod, and A. J. Van Dierendonck, "Applicability of an Augmented GPS for Navigation in the National Airspace System," *Proc. IEEE*, vol. 77, no. 11 (November 1989): 1709–1727.
19. News Item, "Comments on Alternatives to the Microwave Landing System (MLS)," *IEEE Spectrum* (February 1991): 30.
20. T. A. Stansell, Jr., "Civil GPS from a Future Perspective," *Proc. IEEE*, vol. 71, no. 10 (October 1983): 1187–1192.
21. G. L. Burke and A. J. Poggio, "Numerical Electromagnetic Code (NEC)—Method of Moments, Part 1: Program Description—Theory," *Tech. Doc. 116, Naval Electronic Systems Command* (ELEX 3041) (July 1977).
22. A. R. Djordjevic, M. B. Bazdar, G. M. Vitosevic, T. K. Sarkar, and R. F. Harrington, *Analysis of Wire Antennas and Scatterers: Software and User's Manual* (Norwood, Massachusetts: Artech House, 1990).
23. J. Rockway, J. Logan, D. Tam, and S. Li, *The Minninet System: Microcomputer Analysis of Wire Antennas* (Norwood, Massachusetts: Artech House, 1988).
24. J. J. Wang and C. E. Ryan, Jr., "Application of Wire-Grid Modelling to the Design of a Low-Profile Aircraft Antenna," *IEEE Antennas Propagat. Int. Symp. Dig.*, Stanford University (June 1977): 222–225.

25. J. R. Mantz and R. F. Harrington, "Radiation and Scattering from Bodies of Revolution," *J. App. Sci. Res.*, vol. 20 (1969): 405.
26. S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic Scattering by a Surface of Arbitrary Shape," *IEEE Trans. Antennas Propagat.*, vol. AP-30 (May 1982): 409–418.
27. D. L. Knepp and J. Goldhirsh, "Numerical Analysis of Electromagnetic Radiation Properties of Smooth Conducting Bodies of Arbitrary Shape," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (May 1972): 383–388.
28. Edmund K. Miller and Jeremy A. Landt, "Direct Time Domain Techniques for Transient Radiation and Scattering from Wires," *IEEE Proc.*, vol. 68, no. 11 (November 1980): 1396–1423.
29. V. K. Tripp, Engineering Experiment Station, Georgia Institute of Technology, private communication, March 1982.
30. R. C. Hansen (ed.), *Geometric Theory of Diffraction*, selected reprint ser., (New York: IEEE Press, Institute of Electrical and Electronics Engineers, 1981).
31. W. D. Burnside, M. C. Gilreath, R. J. Marhefka, and C. L. Yu, "A Study of DC-135 Aircraft Antenna Patterns," *IEEE Trans. Antenna Propagat.*, vol. AP-23 (May 1975): 309–316.
32. C. L. Yr, W. D. Burnside, and M. C. Gilreath, "Volumetric Pattern Analysis of Airborne Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-26 (September 1978): 636–641.
33. W. P. Cooke and C. E. Ryan, Jr., "A GTD Computer Algorithm for Computing the Radiation Patterns of Aircraft-Mounted Antennas," *IEEE Antennas Propagat. Int. Symp. Prog. Dig.*, Laval University (June 2–6, 1980): 631–634.
34. C. E. Ryan, Jr., "Analysis of Antennas on Finite Circular Cylinders with Conical or Disk End Caps," *IEEE Trans. Antennas Propagat.*, vol. AP-20 (July 1972): 474–476.
35. C. E. Ryan, Jr., and R. Luebbers, "Volumetric Patterns of a Circumferential Slot Antenna on a Conically-Capped Finite Circular Cylinder," ElectroScience Lab., Ohio State Univ. Res. Found. Rep. 2805-3, Cont. DAAA21-69-C-0535, 1970.
36. R. J. Marhefka et al., "Near-Zone Basic Scattering Code User's Manual with Space Station Applications," Technical Report 716199-13, Grant No. NSG 1498.
37. R. J. Marhefka et al., "Antennas on Aircraft, Ships or Any Large Complex Environment," in Y. T. Lo and S. W. Lee (eds.), *Antenna Handbook: Theory, Applications, and Design* (New York: Van Nostrand Reinhold, 1988).

Chapter 41

Radiometer
Antennas

Albin J. Gasiewski

University of Colorado

William F. Croswell

Harris Corporation

Calvin T. Swift

University of Massachusetts

CONTENTS

41.1	INTRODUCTION	41-2
41.2	BASIC PRINCIPLES.....	41-2
41.3	SYSTEM PRINCIPLES.....	41-11
41.4	RADIOMETER ANTENNA TYPES	41-14
41.5	RECENT DEVELOPMENTS AND TRENDS IN RADIOMETRY.....	41-16

41.1 INTRODUCTION

Antennas are widely used as part of passive remote sensing systems for inferring the physical properties of planetary atmospheres and surfaces¹ as well as related applications such as all-weather vision for aviation and imaging through materials for security. Radiometer antennas are similar in type to those used for active systems such as radars and telecommunications links but differ primarily in requiring extreme efficiency and stability. In the case of radar and telecommunications antennas, signal variations of order 0.1 dB or more due to antenna losses and pattern uncertainties may be tolerable. In radiometry the tolerable level of variation imparted by the antenna is often a hundred times (or more) smaller, approaching a level of tens to hundreds of microBels ($1 \mu\text{B} = 10^{-5} \text{ dB}$). This extreme level of stability is required to permit the radiometer system to distinguish thermal fluctuations of order $\sim 0.1 \text{ K}$ or smaller against a background radiation intensity of tens to hundreds of Kelvins.

We can further distinguish between two types of radiometer antennas: those used for Earth remote sensing and those used for radioastronomy. In Earth remote sensing the radiometric system observes a distributed target of large angular extent and relative warmth in temperature. For example, ocean brightness temperatures in the microwave bands are of order 120 K, while land surfaces can be 200-300 K or more.¹ Hence, antenna systems for downward-looking radiometric systems must have small, close-in sidelobes out to the angles where the surface brightness diminishes. In radio astronomy the natural sources (planets, stars, galaxies, and quasars) are of such limited angular extent that they are effectively point sources. In addition, they are immersed in a cold sky of several degrees Kelvin.²⁻³ Techniques such as interferometry are regularly used in radio astronomy to improve resolution at the expense of high sidelobes, and have recently been adopted for remote sensing of atmospheres and surfaces in downward-viewing Earth remote sensing systems.

Owing to the above, radiometer antenna design requires much more precise consideration of antenna performance parameters than most antenna engineers are accustomed to. For example, main beam efficiency, antenna losses, antenna stability, antenna physical temperature, and scanning considerations are extremely important, as are directional gain, sidelobe levels, mismatch, and polarization. In this chapter we include an extended section on basic radiometer principles, along with a description of basic radiometer system types in present or proposed use on spacecraft and aircraft, as well as a discussion of the basic radiometer antenna types commonly in use.

41.2 BASIC PRINCIPLES

For microwave remote sensing applications, the radiometer antenna is used with a very sensitive receiver for detecting and providing a measurement of the electromagnetic radiation intensity received within the solid angle subtended by the antenna beam. The received radiation can originate from a variety of sources, including emission from atmospheric gases and hydrometeors, land, snow, ice, or ocean surface contributions, and the cold cosmic background. The radiation field can undergo scattering from both hydrometeors and the surface. Properties of the atmosphere such as temperature, moisture, and cloud water content influence the radiation intensity within the antenna beam.⁴⁻⁷ Depending upon knowledge of the roughness and dielectric properties of the surface and the relationship of the dielectric properties to the physical properties of the surface, indirect or remote measurements of these properties are feasible.

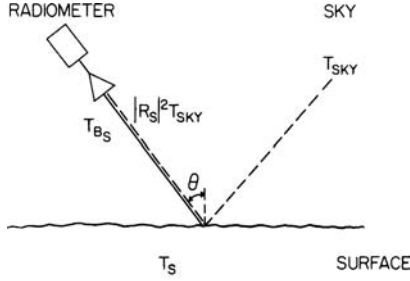


FIGURE 41-1 Typical radiometer observation geometry in microwave remote sensing

The received power detected by the radiometric system is

$$P_{\text{rec}} = \int_{f-\frac{B}{2}}^{f+\frac{B}{2}} \frac{hf}{(e^{hf/kT_A} - 1)} df \approx kT_A B \quad (41-1)$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, $h = 6.63 \times 10^{-34}$ J-s is Planck's constant, B is the radio frequency bandwidth of the receiver in Hz, and T_A is the antenna temperature in Kelvins. The rightmost side of Eq. 41-1 is the Rayleigh-Jeans approximation that holds for frequencies $f \ll kT_A/h$.

Antenna Temperature

The antenna temperature, stated in a manner suppressing polarization and presuming use of the Rayleigh-Jeans approximation,¹ is given by

$$T_A = \frac{\int_0^{2\pi} \int_0^\pi f(\theta, \phi) T_B(\theta, \phi) \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^\pi f(\theta, \phi) \sin \theta d\theta d\phi} \quad (41-2)$$

where $f(\theta, \phi)$ is the radiation pattern of a perfect plane-polarized antenna and $T_B(\theta, \phi)$ is the brightness temperature of the scene observed by a radiometer antenna with the same polarization. A typical scenario for which the antenna is pointed toward a surface is given in Figure 41-1. Here, T_{sky} is the downwelling sky radiation in the specular direction and includes thermal radiation emitted by the atmosphere and from the cosmic background. The brightness temperature observed by the radiometer antenna is the sum of the brightness temperature T_{Bs} emitted by the surface and the energy from the sky radiation scattered by the surface in the specular direction. This brightness temperature $T_B(\theta, \phi)$ is given by

$$T_B(\theta, \phi) = T_{Bs} + |R_s|^2 T_{\text{sky}} \quad (41-3)$$

where $|R_s|^2$ is the reflection coefficient for the air-surface interface and for the polarization of interest.

¹ More precisely, the antenna temperature is derived by integrating the spectral density $\frac{hf}{e^{hf/kT_B(\theta, \phi)} - 1}$ against the antenna pattern over all angles, then inverting this integrated power P according to the function $\frac{hf}{k \ln \left(\frac{hf}{P} + 1 \right)}$.

The emissivity of the surface is given by

$$\epsilon_s = 1 - |R_s|^2 \quad (41-4)$$

and the brightness temperature T_{Bs} is related to the physical surface temperature T_s by the relation

$$T_{Bs} = \epsilon_s T_s \quad (41-5)$$

From Eqs. 41-3, 41-4, and 41-5

$$T_B(\theta, \phi) = (1 - |R_s|^2) T_s + |R_s|^2 T_{sky} \quad (41-6)$$

It is emphasized that these equations do not consider the polarization properties of the antenna and the surface. These properties are discussed in the next subsection.

Polarization

To include the effects of polarization in the expression for antenna temperature, we must consider both the polarization properties of the radiation field and of the antenna itself. For Earth remote sensing, the polarization of the radiation field is described using the Stokes parameters as defined in terms of a simple polarization basis associated with the direction of observation and the surface normal. Consider the coordinate system of Figure 41-2 where the direction of the antenna beam is along (θ, ϕ) relative to a fixed Cartesian coordinate system with the z-axis pointing toward the zenith direction. Let the unit vector \hat{k} point along this direction toward the antenna. Then the unit vectors \hat{v} and \hat{h} are defined as follows:

$$\hat{h} = \frac{\hat{k} \times \hat{z}}{|\hat{k} \times \hat{z}|} \quad \hat{v} = \hat{h} \times \hat{k} \quad (41-7)$$

In general the radiation field propagating along \hat{k} toward the antenna consists of a statistical distribution of plane waves with complex electric field components in both the \hat{v}

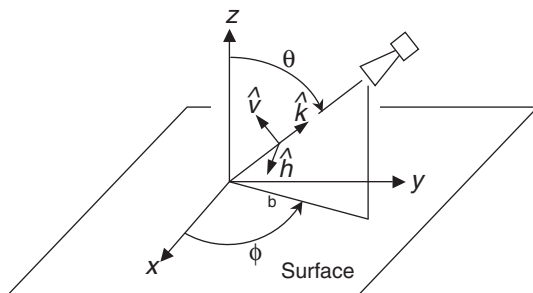


FIGURE 41-2 Polarization of a radiometer antenna with regard to the surface and sky scene

and \hat{h} directions: $\vec{E} = E_v \hat{v} + E_h \hat{h}$. It is possible to fully characterize this distribution of complex field amplitudes using four parameters that describe the powers in the vertical and horizontal polarizations along with the complex correlation between the amplitudes in these two polarizations:

$$\begin{aligned} I &= \frac{1}{2\eta} \left(\langle |E_v|^2 \rangle + \langle |E_h|^2 \rangle \right) \propto T_{BV} + T_{BH} \\ Q &= \frac{1}{2\eta} \left(\langle |E_v|^2 \rangle - \langle |E_h|^2 \rangle \right) \propto T_{BV} - T_{BH} \\ U &= \frac{1}{\eta} \text{Re} \langle E_v E_h^* \rangle \propto T_{BU} \\ V &= \frac{1}{\eta} \text{Im} \langle E_v E_h^* \rangle \propto T_{BV} \end{aligned} \quad (41-8)$$

where $\langle \rangle$ denotes statistical expectation for waves filtered over a small solid angle $d\Omega$ and small frequency interval df , and $\eta = 377\Omega$ is the impedance of free space. These are the four Stokes parameters⁸ for the radiation field. It can be easily shown that $I^2 \leq Q^2 + U^2 + V^2$. Natural surfaces generate upwelling radiation fields with values of I and Q in the range of a hundred Kelvin or more. It has recently been established⁹ that anisotropic surfaces, for example, ocean waves, generate values of U and V of up to a few Kelvins.

To determine the contribution of the four Stokes parameters to the received power it is useful to consider the antenna's gain matrix, defined by analysis of the polarized far-field components of the antenna's radiated field when used as a transmitter:

$$G_{ij}(\theta, \varphi) = \lim_{r \rightarrow \infty} \frac{2\pi r^2 E_i(r, \theta, \varphi) E_j^*(r, \theta, \varphi)}{\eta P_T} \quad (41-9)$$

where P_T is the transmitted power and the subscripts i, j are either \hat{v} (vertical) or \hat{h} (horizontal) field components defined relative to a unit vector in the $-\hat{k}$ direction. The response of the antenna to the Stokes parameters is thus precisely as follows:

$$P_{\text{rec}} = \int_{f-\frac{B}{2}}^{f+\frac{B}{2}} \oint_{4\pi} \left[A_{vv} \left(\frac{I+Q}{2} \right) + \text{Re}\{A_{vh}\}U + \text{Im}\{A_{hv}\}V + A_{hh} \left(\frac{I-Q}{2} \right) \right] d\Omega df \quad (41-10)$$

where $A_{ij}(\theta, \varphi) = \frac{\lambda^2}{4\pi} G_{ij}^*(\theta, \varphi)$ are the effective area matrix elements of the antenna associated with each gain matrix element.¹⁰ Accurate characterization of the antenna gain matrix can be used to properly interpret received power signals in terms of the four Stokes parameters, although good cross polarization isolation (of -30 dB or better) is usually the best way to separate measurements of each of these parameters. Further discussion of this subject is found in later sections of this chapter.^{9,10}

It is useful to note that for many surfaces^{6,11} T_{BH} and T_{BV} have properties such that $T_{BH} - T_{BV} \approx 0$ out to 20° to 30° incidence. Hence, if a narrow-beam low-sidelobe antenna is used for near-nadir observations, the antenna temperature as given in Eq. 41-10 will be largely independent of the polarization of the radiometer antenna. For the antenna designer, the message at this point is rather clear. A radiometer antenna must be designed so that

sidelobes, cross-polarization lobes, and backlobes do not pick up unwanted contributions. Common design parameters for comparing radiometer antennas are the beam efficiency and the cross-polarization index.

Beam Efficiency

The beam efficiency of an antenna is defined as

$$BE(\theta_1) = \frac{\text{power radiated in cone half-angle } \theta_1}{\text{power radiated in } 4\pi \text{ sr}} \quad (41-11)$$

or

$$BE(\theta_1) = \frac{\int_0^{2\pi} \int_0^{\theta_1} f(\theta, \phi) \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) \sin \theta d\theta d\phi} \quad (41-12)$$

It should be noted that polarization is not considered in Eq. 41-12 for simplicity, but will be included later. One can observe from Eqs. 41-11 and 41-12 that to relate one antenna to another in terms of beam efficiency the choice of the angle θ_1 must be made in a standard manner. A common choice for θ_1 is the angle between the main beam axis and first pattern null. Another choice for comparing antennas, such as the wide angle corrugated horn or the multimode horn in which the first null is poorly defined, is to use the criterion for θ_1 as $2\frac{1}{2}$ times the half angle of the 3-dB beamwidth.

The beam efficiency of aperture antennas with non-uniform aperture amplitude distributions has received much attention. Curves comparing one aperture distribution to another are readily available.¹²⁻¹⁴ Examples of the impact of various aperture distributions upon beam efficiency are shown in Figure 41-3 for rectangular and circular apertures as a function of the dimensionless parameter $U = ka \sin \theta$. For uniform circular and rectangular amplitude distributions the first several sidelobes cause pronounced ripples in the beam-efficiency curves, and values of $BE = 95\%$ are not achieved even for large values of U . For more tapered distributions the beam efficiency rises to large values rapidly and independently of aperture shape.

This characteristic of ideal aperture distributions for large beam efficiency is consistent with reduction of wide-angle sidelobes through tapering.^{12,14} Indeed, some antennas such as the multimode, exponential, and corrugated horns have very low sidelobes and backlobes and hence excellent beam efficiencies. Reflector antennas are also sometimes used in radiometry, although in this case the overall beam efficiency is influenced by feed spillover, cross-polarization, and blockage in addition to aperture illumination taper and reflector-surface roughness. These properties will be discussed later on.

The effects of polarization on beam efficiency can be described by the quantity BE_{dp} for the direct (versus cross) polarization:

$$BE_{dp} = \frac{\text{power within angle } \theta_1, \text{ direct polarization}}{\text{total received power, both polarizations}} \quad (41-13)$$

Stated in equation form

$$BE_{dp} = \frac{P_{\theta_1, dp}}{P_{dp} + P_{op}} \quad (41-14)$$

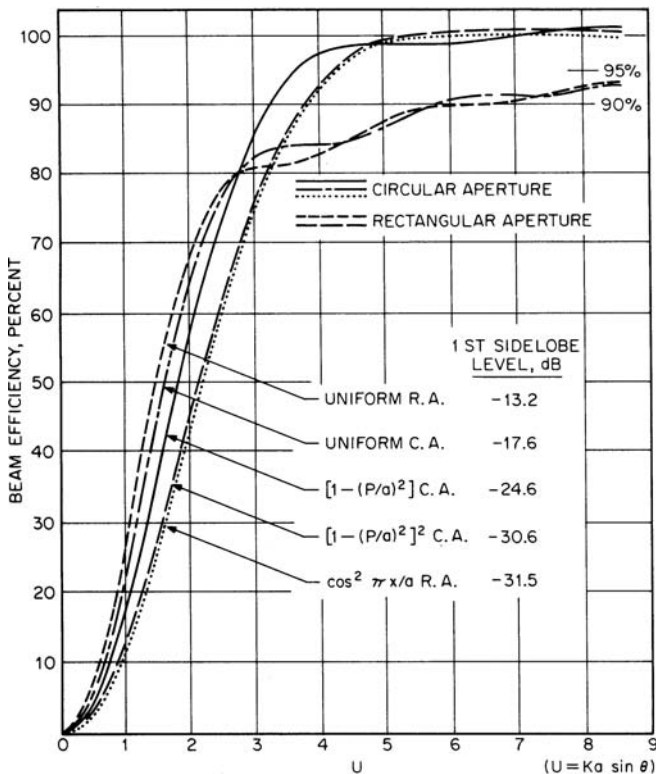


FIGURE 41-3 Comparison of beam efficiencies for rectangular and circular apertures with various aperture distributions

where P_{dp} is the total power received in the direct polarization and P_{op} is the total power received in the orthogonal polarization. The effect of the orthogonal-polarization energy therefore is to decrease the direct-polarization beam efficiency of the antenna. A cross polarization index (CPI) can be defined as¹⁵

$$\text{CPI} = BE_{dp} \frac{P_t}{P_{op}} \quad (41-15)$$

For example, for $BE_{dp} = 85$ percent and on the assumption that the antenna has the property $P_{op}/P_t = 26$ dB, the cross-polarization isolation index is 25 dB.

The significance of cross-polarization levels in a radiometer antenna can be vividly demonstrated by reviewing the earlier discussion on polarization resulting in Eq. 41-10. The brightness temperature of a particular scene is polarization- and angle-dependent. In general, except at near-nadir angles, the brightness temperature of a given surface will be different for vertical and horizontal polarizations. To demonstrate the effect of cross-polarization upon the radiometric measurement of a scene the curves in Figure 41-4 are presented. The bias error ΔT_A in this figure is the error in brightness temperature produced by the integrated cross-polarized lobes observing the surface. This approximation is valid for only narrow-beam antennas in which emission from the surface at a given polarization is constant over $2\frac{1}{2}$ times the 3-dB beamwidths

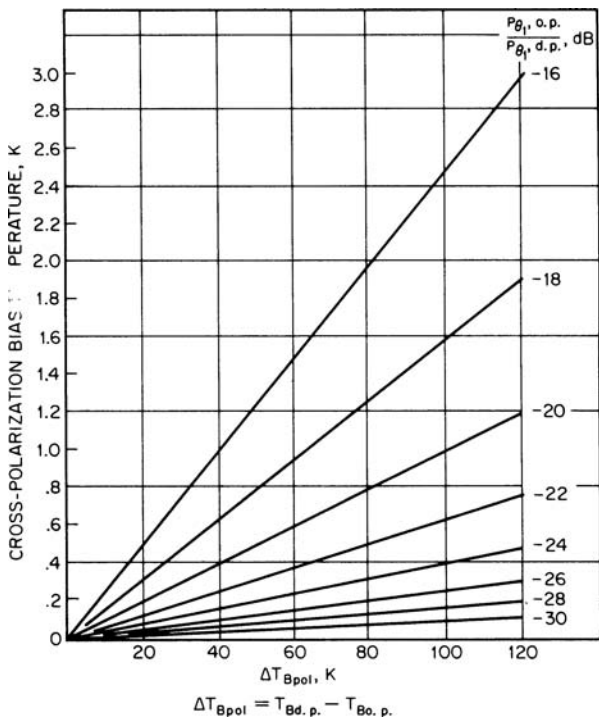


FIGURE 41-4 Effect of antenna cross-polarization power on radiometer bias as a function of scene polarization difference temperature

of the antenna. For precision measurements of ocean temperature in which accuracies of 0.3 K or better are desired, values of cross-polarization energy less than 28 dB are required.

Stray Radiation

For narrow-beam antennas, $T_B(\theta, \phi)$ is relatively constant over the main beam, and Eq. 41-2 can be written as

$$T_A = T_B(\theta, \phi)BE_{\theta_1} + \frac{\int_0^{2\pi} \int_{\theta_1}^{\pi} f(\theta, \phi)T_B(\theta, \phi) \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi} f(\theta, \phi) \sin \theta d\theta d\phi} \quad (41-16)$$

If the assumption that $T_B(\theta, \phi)$ is constant in the second term of Eq. 41-16 is made, this equation may be written as

$$T_A = T_B(\theta, \phi)BE_{\theta_1} + T_B(\theta, \phi)[1 - BE_{\theta_1}] \quad (41-17)$$

The second term in Eq. 41-17 is the so-called stray radiation contribution. For radioastronomy and some Earth remote sensing applications, the near-in and relatively far-out sidelobes observe cold sky surrounding the main beam, thus justifying the assumptions made in Eq. 41-17. For radioastronomy applications the design emphasis must be placed

on minimizing contributions from backlobes which point toward the warm earth. For Earth remote sensing applications, the close-in and wide-angle sidelobes observe brightness temperatures in the same range as the main beam (100–270 K), while in many configurations the backlobes point to the cold sky. Concepts such as stray radiation are useful in that they provide a bound on the ultimate accuracy of the T_A measurement. Simplistic forms to compute stray radiation such as Eq. 41-17 should be avoided unless great care has been exercised to understand the amplitude and angular distribution of brightness temperatures in the scene.

Ohmic and Reflection Losses

The ohmic losses in the antenna will modify the apparent temperature observed by the radiometric system by the relation⁶

$$T_a = (1 - \epsilon_L) T_A + \epsilon_L T_0 \quad (41-18)$$

where T_a is the apparent antenna temperature of a source whose antenna temperature (observed without loss) is T_A , T_0 is the physical temperature of the antenna, and ϵ_L represents the fractional power loss in the antenna.

The significance of physical losses in the antenna upon the absolute accuracy of remotely sensed surface properties can be seen using the following example. Assume that the antenna physical temperature is that of room temperature, for example, $T_0 \approx 300$ K. Typical ocean and land scenes exhibit (lossless) antenna temperatures between 100 and 300 K. Using this information, the effect of antenna losses upon measurement accuracy is given in Figure 41-5. For some ocean-temperature applications ($T_A \approx 102$ K), absolute

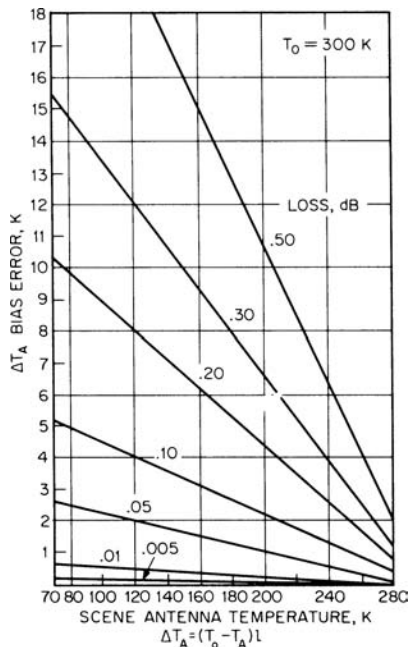


FIGURE 41-5 Effects of losses upon scene antenna temperature

accuracies of 0.3 K or smaller are of importance. Hence, very small losses (≤ 0.005 dB) are of significance. For antennas in which losses are significant, this loss must be stable, well characterized, and the physical temperature of the antenna known to high precision so that the added emission can be treated as a known bias. For example, a Potter horn built for a precision S-band radiometer⁶ exhibited an acceptably stable loss of ≤ 0.1 dB. Studies of corrugated and multimode horns¹⁶ indicate small but significant losses when these are used as radiometric antennas. To ensure stability in calibration, antennas may be enclosed in thermally stable boxes with very low loss radomes.

The effect of mismatch at the input to the radiometer upon the observed antenna temperature can be expressed as⁶

$$T_a = (1 - \rho)[(1 - \epsilon_L)T_A + \epsilon_L T_0] + \rho T_R \quad (41-19)$$

where ρ is the reflection coefficient of the antenna and T_R is the microwave temperature seen looking into the receiver. The effects of small mismatches are very important for precision measurements. Stability of this mismatch will allow one to treat this error as a bias. Again, as in the case of physical loss, the thermal stability of the antenna impedance is important.

Calibration

Calibration is performed by stimulating the input of a radiometer with known temperatures and then identifying the system response parameters. The gain and offset of a radiometer system are subject to significant drift compared to the typical required brightness precision. Accordingly, calibration of the entire radiometer system, including the antenna, is desired to compensate for such variations. Three common methods of calibration are shown in Figure 41-6. The steerable-antenna method (see Figure 41-6a) requires that the radiometer antenna be able to be periodically pointed toward a cold reference such as space (ideally the galactic pole) to provide a measurement of system gain. The antenna can also be pointed at a warm reference, such as a thermally stable lake, in order to obtain a second calibration point for determination of both gain and offset. References can also be microwave blackbody targets at distinct temperatures, for example, a porous microwave absorber located in a container filled with liquid nitrogen. This type of load is commonly used in laboratory calibration and for pre-flight calibration of airborne radiometers, provided that the radiometers have been designed to be very stable over long periods of time. The emission characteristics of such blackbody targets must be well understood.¹⁷ Accurate calibration requires that the references have temperatures that bound the range of variation of scene antenna temperature.

For applications where the antenna cannot be steered toward references, a switchable two-antenna technique (see Figure 41-6b) can be used although variations in system response due to differences between the two antennas must be determined. Another disadvantage of this scheme are switch losses that might change as a function of time, depending upon the switch technology and method

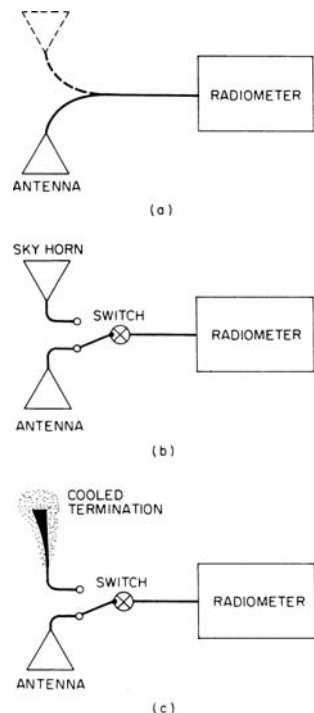


FIGURE 41-6 Radiometer calibration methods: (a) steerable antenna, (b) reference antenna and switch, and (c) cooled termination and switch

of thermal control. The third method (see Figure 41-6c) uses cooled terminations and has the advantage of not requiring a separate or movable antenna. This technique is very useful for systems in which rapid internal calibration is necessary.¹⁸ Cooled terminations can be fabricated either using Peltier effect coolers or by connecting to the input of a low noise amplifier. Use of two switches and a pair of terminations at two distinct temperatures permits determination of both gain and offset. Of course, switch losses are common to the methods in Figure 41-6b and c.

Calibration must be performed often enough to identify significant variations in gain and offset. The time stability of the radiometer components, which depends to a great extent on thermal stability, determines the required maximum calibration period. End-to-end calibration of fully polarimetric radiometers has been demonstrated.^{19,20}

41.3 SYSTEM PRINCIPLES

Radiometer Types

A variety of radiometer architectures have been used for radio astronomy and Earth remote sensing (for example, Hidy et al⁶). Four basic types commonly in use are described in simplistic block-diagram form in Figure 41-7.

The simplest of these is the total-power radiometer shown in Figure 41-7a. The output voltage of this radiometer, which is not modulated, can be expressed as

$$V = G(T_A + T_N) \quad (41-20)$$

where T_A is the antenna temperature, T_N is the system noise temperature referred to the antenna input terminals, and G is the power gain of the receiver system. The total power radiometer can resolve a change in scene temperature in a dynamic system of

$$\Delta T = \frac{T_B + T_R}{\sqrt{B\tau}} \quad (41-21)$$

where

- T_B = brightness temperature of the scene, K
- T_R = effective receiver noise temperature, K
- B = predetection bandwidth, Hz
- τ = integration time, s

The absolute accuracy of this system is dependent upon the stability of the radiometer after calibration using hot and cold references. Use of state-of-the-art receiver technology and precision thermal regulation can usually result in accuracies of better than ~1 K over reasonable periods of time. For spaceborne radiometers the calibration of this system is improved by spinning the spacecraft antenna so that cold sky is observed during each revolution, as in the calibration method of Figure 41-6a. This spinning-antenna system is equivalent to the signal-modulated Dicke-switched radiometer discussed next, except that the switching circulator is eliminated and the modulation frequency is reduced.²¹

Another commonly used radiometer type is the Dicke switched radiometer²² of Figure 41-7b. The advantage of the Dicke radiometer is that variability of the noise from the receiver T_N is eliminated so that the output voltage is given by

$$V = G[T_A(1 - \epsilon_s) - T_{\text{ref}}(1 - \epsilon_R) + (\epsilon_s - \epsilon_R)T_0] \quad (41-22)$$

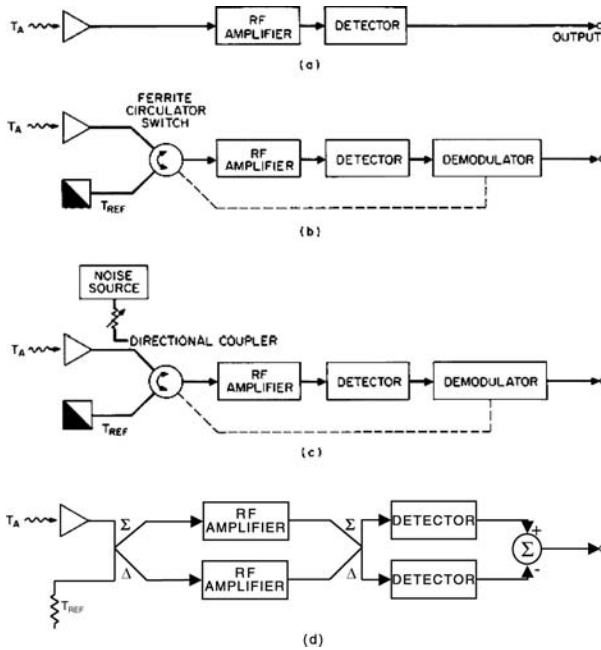


FIGURE 41-7 Simplified diagram of basic radiometer types (after Hidy et al⁶): (a) Total power radiometer; (b) Dicke switched radiometer; (c) Noise-injection switched radiometer; (d) Differential radiometer

where ϵ_s and ϵ_r are the ohmic losses in the signal and reference arms and T_0 is the physical temperature of the radiometer components. Provided that T_{ref} is maintained close to T_A , the system gain variations have less of an impact on overall accuracy than for a total power system. However, the rms integration noise is twice as large. A variation on the Dicke switched radiometer is the noise injection signal-modulated type shown in simplified form in Figure 41-7c. The noise injection method eliminates gain instability by permitting noise power to be added or subtracted to the reference input of the radiometer based on negative feedback. Various methods of performing this variable noise injection include use of variable calibrated attenuators and pulse-width noise switching. Variable-height pulse modulation of a noise diode feedback circuit can also be employed.⁶ A digital version of the noise-injection modulated radiometer has been devised and analyzed²³ and implemented.²⁴

A fourth type of radiometer is the differential radiometer²⁵ shown in Figure 41-7d. This system is based on a sum-difference hybrid combiner, dual amplifier chains, and a correlator stage formed from a second sum-difference combiner. It has the decided advantages of a differential response as in Eq. 41-22 but also that amplifier noise and gain fluctuation noise are reduced in the output. However, it requires two amplifier and detector chains, and is thus more expensive. For the differential radiometer the integration noise is $\sqrt{2}$ times that of the total power radiometer.

Radiometric Imaging Systems

The most common radiometer system in Earth remote sensing is the cross-track imaging type depicted in Figure 41-8a. Imaging on a spacecraft or aircraft is usually achieved by

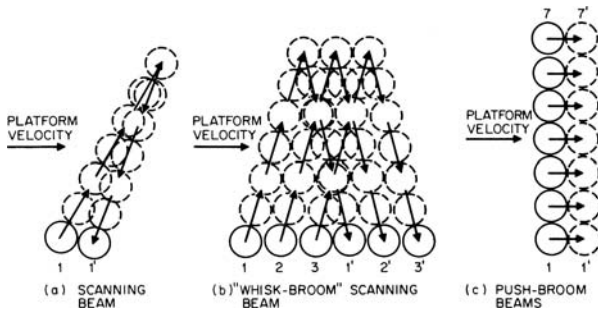


FIGURE 41-8 Ground spot patterns of cross-track scanned radiometer-antenna systems

uni- or bidirectional scanning of a reflector/antenna/feed system relative to a stable moving platform. Imaging radiometers provide wide-swath coverage of the surface or atmospheric scene. The spatial resolution of the scanning system is limited by antenna size, which in turn is limited by the size of aircraft ports or launch vehicle fairings. When finer spatial resolution is required, large deployable antennas that are limited in scan rate and required to spin with the radiometers as a single unit can be employed. Owing to the combination of scan rate and spacecraft velocity, a single-beam radiometer antenna of the scanning type may not produce contiguous resolution cells on the surface at the beginning and end of each rotation, and may offer only a small amount of integration time per sample. To produce contiguous coverage, a whiskbroom radiometer antenna system can be employed as in Figure 41-8b. Here each beam in the whiskbroom requires an independent antenna port and radiometer.²⁶ The beam patterns can be produced by multiple feeds in a single reflector system.

As even larger radiometer antennas are required to obtain better surface resolution, physical motion to obtain wide swath coverage may not be possible. To provide swath-width coverage without mechanical scanning a pushbroom radiometer system can be used, as in Figure 41-8c. In this design an independent radiometer is connected to each of several antennas with adjacent beams. Regular sampling of the pushbroom beam produces a surface radiometric brightness temperature map.

A mechanically scanned beam can be implemented by pointing a feed horn in the direction of platform motion and rotating a reflector tilted at 45° to this axis about this axis. This simple cross-track scanned system is used in a number of airborne and low-Earth orbiting (LEO) spaceborne systems, for example the NOAA Advanced Microwave Sounding Unit (AMSU).²⁷ Reference sources can be readily viewed during the scan cycle for calibration. Two disadvantages are that the surface spot size changes throughout the scan (being smaller at nadir) and the polarization basis of the feedhorn rotates throughout the scan. A conical scanning system resolves these problems by providing constant incidence angle and fixed polarization basis throughout the scan, but requires the radiometer to rotate with the reflector antenna. A scheme for electronically rotating the polarization basis of a radiometer has been demonstrated to permit the radiometer to be fixed to the platform while maintaining polarization alignment with the Earth throughout the scan.⁹ Airborne systems have been built to provide both conical and cross-track scanning.²⁸

Because of requirements to distinguish signals from parameters such as salinity and temperature for the ocean, or clouds and water vapor for the atmosphere, multiple frequencies are commonly required to be integrated into scanning,¹⁶ whiskbroom, or pushbroom radiometric systems. Such multifrequency systems are usually designed as multiple feeds with a single main reflector. Combining frequencies in a single broadband corrugated horn is feasible over nearly 2:1 bands with low-loss properties and also for

bands of approximately odd harmonics (for example, 3:1, 5:1). A very broadband horn with over 5:1 bandwidth has been designed and used in space with a scanning system.²⁹ This antenna, however, exhibits large losses which must be compensated as large bias errors.

41.4 RADIOMETER ANTENNA TYPES

Horns

A straightforward design for a moderate-beamwidth (10° to 30°) radiometer antenna is the electromagnetic horn; however, because of the requirement for high beam efficiency (and, therefore, for low sidelobes) specialized horn designs are necessary for radiometer applications. Radiometer horn antennas are designed so that the normally high E -plane sidelobes are reduced to an acceptable level by aperture field tapering. The H -plane sidelobes (typically -23 dB) are already low enough to achieve a sufficiently high beam efficiency for moderate beamwidths, although they can be further reduced by tapering. Tapering can be accomplished using several techniques. The simplest is to utilize multiple modes³⁰ in the radiating-horn aperture to provide sidelobe cancellation in the far-field pattern. The higher-order modes are excited by a step or sudden change in the cross section of the waveguide or horn taper. The bandwidth of a multimode or dual-mode horn is limited to a few percent because of the difference between the modal phase velocities in the tapered waveguide horn section. However, a shortened version³¹ of the Potter horn can exhibit good pattern characteristics over a bandwidth approaching 10 percent. The performances of the square multimode horn and the conical dual-mode horn are quite similar.

For radiometer applications requiring a larger bandwidth and low sidelobes the class of corrugated horns³² is attractive. As with the multimode-type horn, the purpose of the corrugated-horn design is to reduce the normally high E -plane sidelobes by tapering. This is accomplished by designing the corrugations so as to decrease the current along the corrugated wall, thus producing a tapered (approximately cosine) E -plane aperture field distribution. This tapered distribution can be maintained over a bandwidth approaching 2:1 but is usually limited to the operating bandwidth of the feed waveguide. Multiple bands of odd harmonic order can be implemented using the same corrugations.

Another type of horn which shows promise as a wideband radiometer antenna is the exponential horn with a specially flared aperture.³³ A similar wideband horn is one which is flared like a trumpet.³⁴ The design approach of these wideband horns is to eliminate sharp discontinuities and provide a smooth transition between the horn modes and free space.

Beam-efficiency calculations are given in Figure 41-9 for a circular aperture with a radial aperture distribution equivalent to that of the TE_{11} -mode H -plane distribution and with a quadratic phase taper, as being representative of the beam efficiencies obtained from radiometer horns. The figure also illustrates that, in order to achieve high beam efficiencies, the horn must be designed for small phase taper, either by decreasing the horn flare angle or by using a phase-correcting aperture lens.

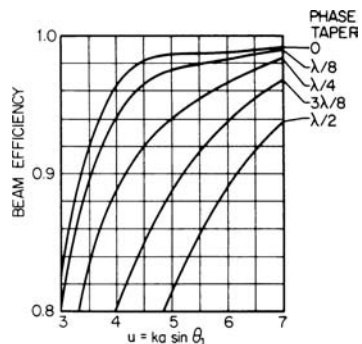


FIGURE 41-9 Effect of aperture phase taper on beam efficiency for a conical corrugated or dual-mode horn

Feed-Reflector Systems

For radiometer applications requiring narrow beamwidths or multiple bands, reflector antennas are more appropriate since highly tapered reflector illuminations can yield the much lower sidelobes necessary for high beam efficiency. The feed elements should be designed for spillover minimization, which represents additional beam-efficiency degradation. Aperture efficiency is generally a secondary concern for radiometers. The reflector surface must be constructed precisely (usually a machined and possibly polished surface is required) to minimize loss in beam efficiency due to statistical roughness.

The overall beam efficiency for a reflector antenna can be expressed as

$$BE = BE_s BE_f BE_\delta \quad (41-23)$$

where BE_s is the main beam efficiency obtained by integration of the reflector secondary pattern, neglecting the back radiation, BE_f is the spillover beam efficiency of the feed evaluated at the angle of the edge illumination, and BE_δ is the beam efficiency due to reflector surface roughness. Calculations of beam efficiencies BE_s for a circular aperture with raised quadratic and raised quadratic squared distributions are presented in Figures 41-10 and 41-11 and representative of reflector illuminations with edge tapers of -20 , -10.5 , -8 and -6 dB. The dashed curve for uniform illumination is included for reference. Figure 41-12 shows the effect of edge taper on the spillover efficiency BE_f . The calculations are for a conical dual-mode horn with no phase error as being typical of feed patterns used in horn-reflector radiometer antennas. The reduction in beam efficiency BE_δ due to reflector-surface roughness can be obtained from the analysis of Ruze³⁵ as

$$BE_\delta = BE_s \exp(-\delta^2) + \Delta BE \quad (41-24)$$

where $\delta = 4\pi\sigma/\lambda$, σ is the rms surface roughness, and ΔBE is a correction term which accounts for the nonzero correlation length c of the surface error. Specifically,

$$\Delta BE = \exp(-\delta^2) \sum_{n=1}^{\infty} (\delta^{2n}/n!) [1 - \exp(-(uc/D)^2/n)] \quad (41-25)$$

where D is the antenna diameter and $u = ka \sin \theta_l$ with θ_l defined as in Eq. 41-11. Figure 41-13 is a plot of the reduction factor versus rms surface roughness and correlation length. It should

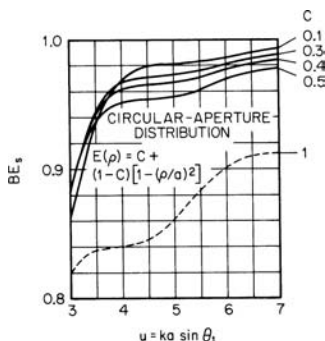


FIGURE 41-10 Beam efficiency for a circular aperture with parabola-on-a-pedestal distribution

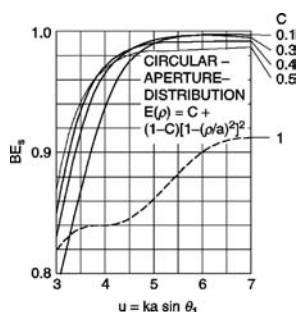


FIGURE 41-11 Beam efficiency for a circular aperture with a parabola-on-a-pedestal-squared distribution

be noted from Eq. 41-24 that for beam efficiencies greater than 90 percent the reflector surface should be $\lambda/50$ or smoother.

An additional concern for offset reflectors is the cross-polarization level, which can complicate the interpretation of antenna temperature data as discussed earlier. The cross-polarization for reflectors decreases for larger focal length to diameter ratios and smaller offsets;³⁶ however, these parameters must be optimized to minimize feed and spar blockage effects. The BE_s curves in Figures 41-10 and 41-11 neglect the effects of feed and strut blockage. Indeed, these effects cannot be treated simply as an aperture blockage as in gain calculations, and the secondary pattern beam efficiency BE_s therefore should be recomputed with the wide-angle scattering from feed and spars included.

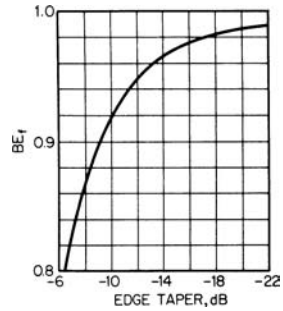


FIGURE 41-12 Feed beam efficiency factor for a reflector antenna

Phased Arrays

Phased arrays have been successfully employed as radiometer antennas^{37,38} when beam scanning is required or volume constraints indicate that an array is the appropriate antenna type. High beam efficiency is obtained through amplitude tapering within the feed distribution network. In the design of feed networks for radiometer array antennas, internal line losses must be minimized. If line losses are not excessive, stabilization of the losses through temperature control can be used in combination with calibration techniques to correct the radiometric temperature data. Resistors and terminations within the feed network should be avoided or used with discretion, since noise emitted by such components into the receiver could desensitize and confound the radiometric measurement.

Microstrip Patch Radiometer Antennas

Although the ohmic efficiency of microstrip antennas is not high, arrays of microstrip patch antenna elements are attractive for airborne radiometry due to their potential for conforming to arbitrary surface shapes.³⁹ Accordingly, it can be expected that as feed and coupling techniques for improving main beam efficiency and thermal measurement and compensation techniques improve, the use of microstrip antennas in radiometry will become more widespread.

41.5 RECENT DEVELOPMENTS AND TRENDS IN RADIOMETRY

Overview

Several major developments in radiometry and radiometer antennas have occurred in the areas of

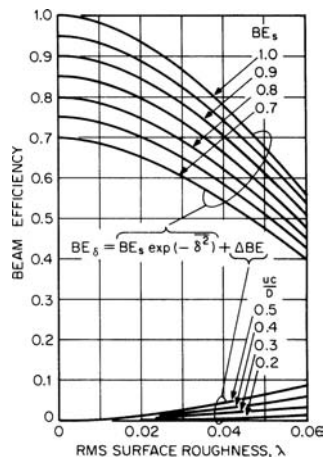


FIGURE 41-13 Beam-efficiency factor of a random rough surface reflector

precision radiometric imaging and sounding and interferometry for Earth applications. Two particularly important applications of radiometry are polarimetric imaging of the Earth for ocean wind direction sensing and geostationary microwave imaging and sounding. These radiometric techniques rely on state-of-the-art radiometer antennas or systems of antennas.

Spaceborne Imaging Radiometers

Since the launch of the Scanning Multichannel Microwave Radiometer (SMMR)²⁹ system in 1978, the scientific and operational value of microwave imagery for measuring Earth surface parameters such as soil moisture, sea surface temperature, snow water equivalent, sea ice concentration, and atmospheric parameters such as precipitable water vapor, rain rate, and cloud liquid water has been well established. Conical microwave imaging is performed using a rotating offset parabolic reflector antenna with $\sim 55^\circ$ incidence angle scanned about a vertical axis. The resulting conical scan provides a polarization basis that coincides with the Earth's natural basis⁹ provided that the feedhorn and radiometers also rotate with the scan. Examples of such systems include SMMR, the DMSP Special Sensor Microwave/Imager (SSM/I),⁴⁰ the JAXA Advanced Microwave Scanning Radiometer (AMSR-E) on board the NASA EOS Aqua spacecraft,⁴¹ the Tropical Rainfall Measurement Mission (TRMM) Microwave Imager,⁴² and the U.S. Navy's WindSat polarimetric imaging sensor.⁴³ These systems use either closely spaced arrays of multi-polarization feedhorns or multiband multipolarization feedhorns to achieve nearly coincident beams at several frequency bands. The selected bands are located at key transmission window or absorption line frequencies that collectively span over a decade in bandwidth. These sensors provide dual polarization imagery, while the WindSat sensor also has provided the first fully polarimetric (four Stokes parameter) imagery from space.

The SSM/I system consists of a seven-channel, four-frequency, linearly polarized microwave radiometer system operating at 19.3, 22.2, 37.0, and 85.5 GHz.⁴⁰ This system was launched into low Earth orbit and travels at a velocity of 6.58 km/s at an altitude of 833 km. The SSM/I Earth coverage in a single day is shown in Figure 41-15. The scanning antenna depicted in Figure 41-14 is an offset reflector fed by a multifrequency corrugated horn.

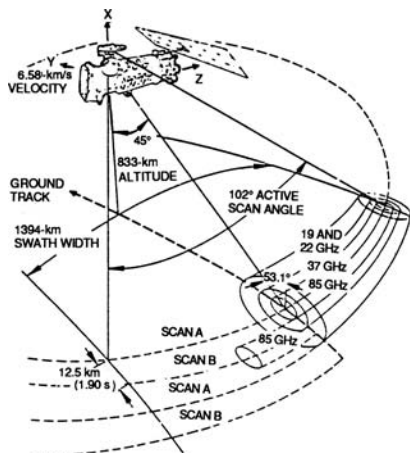


FIGURE 41-14 SSM/I scan geometry (after Hollinger et al⁴⁴ © IEEE 1990)

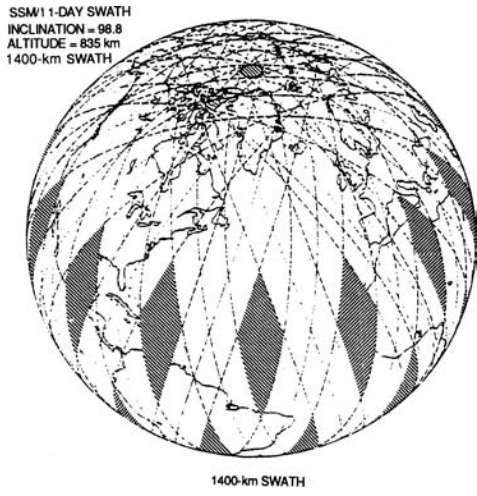


FIGURE 41-15 SSM/I Earth coverage of SSM/I in a 24-hour period. Shaded areas are not observed in this time period (after Hollinger et al⁴⁴ © IEEE 1990).

A similar corrugated horn at low-frequency bands was used on the SMMR system. The SSM/I feed horn has a flare angle of 58° , giving an equivalent edge taper of about 17 dB for all frequencies. The SMMR horn (Figure 41-16) used a set of ring-loaded corrugations. This horn design has been modified to include straight-sided grooves of alternating depth. Multifrequency performance is achieved by coupling waveguides at various points along the feed waveguide and/or horn body. Filters were placed in the various waveguide ports to minimize multiband coupling. These remarkable feedhorns were designed by Dr. K. A. Green and have significant losses of 0.5 to 0.7 dB, depending on operating band.

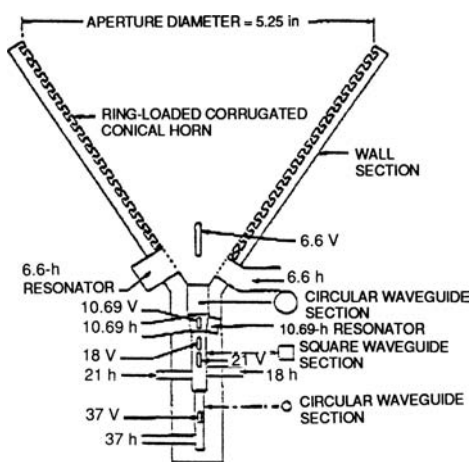


FIGURE 41-16 Multiband corrugated feed horn used for SMMR (after Njoku et al²¹ © IEEE 1980)

This loss must be calibrated and removed from the antenna temperature as a bias (as in Figure 41-5). Cross-track scanning radiometers such as the NOAA Advanced Microwave Sounding Unit (AMSU²⁷) use simpler scanning antenna configurations consisting of feedhorns directed toward flat 45° tilted rotating mirrors. The spatial resolution varies throughout the scan, as does the polarization vector.

Synthetic Aperture Interferometry

Direct imaging radiometers such as AMSR-E and WindSat map brightness temperatures by scanning their antennas mechanically across the field of view. The resolution of these filled-aperture radiometers for a particular channel is determined by the size of the main reflector and wavelength. In contrast, interferometric imaging radiometers generate brightness images indirectly by measuring the Fourier transform of the brightness temperature distribution over a wide field of view. Interferometry is implemented by cross-correlating signals received from a number of small broad-beam antennas that are spatially separated so as to achieve a near-uniform distributions of baselines. Each of these antennas have overlapping individual fields of view. An example of a pair of elements is given in Figure 41-17.⁴⁵ A generic antenna correlator is depicted in Figure 41-18. Because the effective resolution of the system depends on the maximum baseline dimensions rather than the antenna element size, the system is referred to as having a synthetic (as opposed to real) aperture. Synthetic aperture interferometry has been used for decades⁴⁶ in radioastronomy at frequencies of up to 86 GHz for terrestrial systems⁴⁷ and baselines of up to 2.6 Earth diameters for combined ground- and space-based receivers.⁴⁸ The first spaceborne synthetic aperture interferometer capable of mapping global Soil Moisture and Ocean Salinity (SMOS) is being designed by the European Space Agency (see Figure 41-19).^{49,50} SMOS implements an L-band synthetic aperture imager using 69 patch antenna elements in a Y-shaped array with a maximum baseline of 6.75 m with sensitivity over 1400–1427 MHz.

For the synthetic aperture radiometer, the scene temperature sensitivity is given by

$$\Delta T = \sqrt{\frac{(T_0 + T_{R1})(T_0 + T_{R2})}{B\tau} \frac{(2N_x + 1)(2N_y + 1)}{2}} \quad (41-26)$$

where T_{R1} and T_{R2} are the receiver noises of the two receiver channels and N_x and N_y are the numbers of antenna elements in the X and Y planes. The standard deviation in the noise floor of the image has been increased by the root sum square of the noise in individual measurements. An extension of the correlation technique can be applied to a high-resolution thinned array.^{51,52}

An airborne thinned array synthetic aperture imaging radiometer operating at X-band has been demonstrated in preparation for the NASA Global Precipitation Mission.⁵³ Other applications of correlation radiometry include measurements of the third and fourth Stokes parameters⁵⁴ and emission spectra for atmospheric profiling applications.⁴⁵

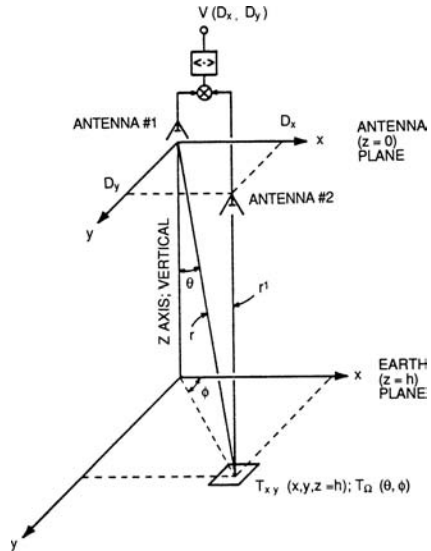


FIGURE 41-17 A two-dimensional Earth-viewing interferometer. The two antennas in the $z = 0$ plane have antenna patterns directed toward Earth (after Ruf et al⁴⁵ © IEEE 1988)

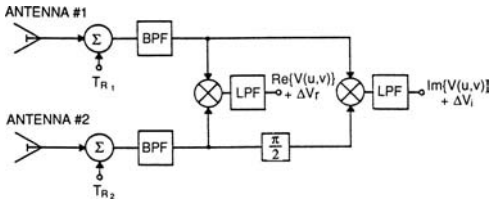


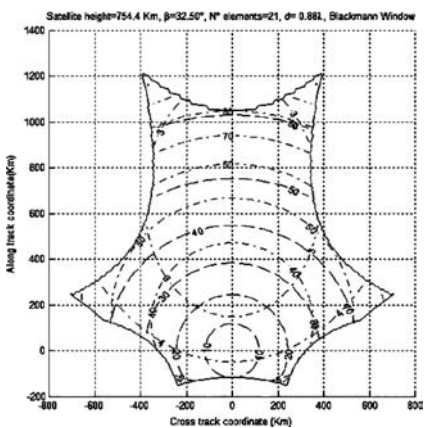
FIGURE 41-18 Quadrature (complex) antenna correlator, where receiver noise is added at the antenna terminals

Geostationary Microwave Imaging and Sounding

Microwave imaging of the Earth from geostationary orbit offer the unique possibility of tracking storms, weather fronts, and hurricane rainbands that are hidden to infrared and optical sensors by clouds and at the time scales at which these weather systems naturally evolve (~15 minutes or less). However, the high orbital altitude presents unique engineering challenges in designing an efficient multiband radiometer antenna with extremely narrow beams and a low mass-and-power scanning mechanism. One such design utilizes a 2-meter diameter main parabolic reflector and moveable Cassegrain subreflector to provide a rapid scan of its beams over a portion of the Earth's disk.⁵⁵ The entire antenna slews at a slow rate so as to permit imaging over the entire disk with little momentum transfer to the host platform. The five-band GEostationary Microwave (GEM) radiometer provides a 3-dB subsatellite spatial resolution of ~16 km at the highest of its sounding frequencies (424 GHz), and progressively lower resolution but higher radiometric sensitivity at its lower frequencies (380, 183, 119, and 55 GHz). To maintain high main beam efficiency, the subreflector is limited to scanning a range of ~0.3° around the antenna boresite. Geostationary imaging of the earth in microwave bands is also being studied using synthetic aperture interferometry⁵⁶ with the potential advantage of implementing a system having no moving parts.



(a)



(b)

FIGURE 41-19 (a) Rendering of the SMOS satellite showing its Y-array of antenna elements directed toward Earth and (b) synthesized SMOS field of view on the Earth's surface (after Font et al.⁵⁰ © IEEE 2004)

REFERENCES

1. D. H. Staelin and P. W. Rosenkranz, "High Resolution Passive Microwave Satellites," *MIT Res. Lab. Electronics* (April 1977).
2. P. W. Bounton, R. A. Stokes, and D. T. Wilkinson, "Primeval Fireball at $\lambda = 3$ mm," *Phys. Rev. Lett.*, vol. 21 (1968): 462.
3. J. R. Shakeshaft and A. S. Webster, "Microwave Background in Steady State Universe," *Nature*, vol. 217 (1968): 339.
4. A. J. Gasiewski, "Microwave Radiative Transfer in Hydrometeors," *Atmospheric Remote Sensing by Microwave Radiometry*, Chap. 3, M. A. Janssen, (ed.) (New York: John Wiley and Sons, 1993).
5. W. H. Peak, "The Microwave Radiometer as a Remote Sensing Instrument," *Electrosci. Lab. Rep.* 1907-8, Ohio State University, Columbus, Ohio, January 17, 1969.
6. G. M. Hidy, W. F. Hall, W. N. Hardy, W. W. Ho, A. C. Jones, A. W. Love, J. Van Melle, H. H. Wang, and A. E. Wheeler, "Development of a Satellite Microwave Radiometer to Sense the Surface Temperature of the World Oceans," NASA CR-1060, National Aeronautics and Space Administration, Washington, February 1972.
7. C. T. Swift, "Passive Microwave Remote Sensing of the Ocean—A Review," *Boundary-Layer Meteorology*, vol. 18 (1980): 25–54.
8. L. Tsang, J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing* (New York: John Wiley and Sons, 1985).
9. A. J. Gasiewski and D. B. Kunkee, "Polarized Microwave Thermal Emission from Water Waves," *Radio Science*, vol. 29, no. 6 (November–December, 1994): 1449–1466.
10. L. C. H. Papas, *Theory of Electromagnetic Wave Propagation* (New York: Dover Publications, Inc., 1988).
11. F. B. Beck, "Antenna Pattern Corrections to Microwave Radiometer Temperature Calculations," *Radio Sci.*, vol. 10, no. 10 (October 1975): 839–845.
12. A. F. Sciambi, "The Effect of the Aperture Illumination on the Circular Aperture Antenna Pattern Characteristics," *Microwave J.*, (August 1965): 79–81.
13. R. T. Nash, "Beam Efficiency Limitations of Large Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-12 (November 1964): 691–694.
14. J. Ruze, "Circular Aperture Synthesis," *IEEE Trans. Antennas Propagat.*, vol. AP-12 (November 1964): 691–694.
15. W. H. Kummer, A. T. Villeneuve, and A. F. Seaton, "Advanced Microwave Radiometer Antenna System Study," NASA Cont. NAS 5-20738, Hughes Aircraft Company, Antenna Department, Culver City, CA, August 1976.
16. R. Caldecott, C. A. Mentzer, L. Peters, and J. Toth, "High Performance S-Band Horn Antennas for Radiometer Use," NASA CR-2133, National Aeronautics and Space Administration, Washington, January 1973.
17. D. M. Jackson, "Calibration of Millimeter and Submillimeter-Wave Radiometers with Application to Clear-Air Remote Sensing of the Troposphere," Ph.D. thesis, Georgia Institute of Technology, July 1999.
18. I. Corbella, A. J. Gasiewski, M. Klein, V. Leuski, A. J. Francavilla, and J. R. Piepmeier, "On-board Accurate Calibration of Dual-Channel Radiometers Using Internal and External References," *IEEE Trans. Microwave Theory Tech.*, vol. 50, no. 7 (July 2002): 1816–1820.
19. A. J. Gasiewski and D. B. Kunkee, "Calibration and Applications of Polarization Correlating Radiometers," *IEEE Trans. Microwave Theory Tech.*, vol. 41, no. 5 (May 1993): 767–773.
20. J. Lahtinen, A. J. Gasiewski, M. Klein, and I. Corbella, "A Calibration Method for Fully Polarimetric Radiometers," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 3 (March 2003): 588–602.
21. E. G. Njoku, J. M. Stacey, and F. T. Banath, "The SeaSat Scanning Multichannel Microwave Radiometer (SMMR): Instrument Description and Performance," *IEEE J. Oceanic Eng.*, vol. OE-5, no. 2 (April 1980): 100–115.

22. R. H. Dicke, "The Measurement of Thermal Radiation at Microwave Frequencies," *Rev. Sci. Instrumen.*, vol. 17 (1946): 268–275.
23. W. D. Stanley, "Digital Simulation of Dynamic Processes in Radiometer Systems," Final Rep., NASA Cont. NASI-14193 No. 46, Old Dominion University, Norfolk, VA, May 1980.
24. R. W. Lawrence, "An Investigation of Radiometer Design Using Digital Processing Technique," Master's thesis, Old Dominion University, Norfolk, VA, June 1981.
25. N. Jarosik, C. L. Bennett, M. Halpern, G. Hinshaw, A. Kogut, M. Limon, S. S. Meyer, L. Page, M. Pospieszalski, D. N. Spergel, G. S. Tucker, D.T. Wilkinson, E. Wollack, E. L. Wright, and Z. Zhang, "Design, Implementation and Testing of the MAP Radiometers," *Astrophys. J. Supplement*, vol. 145, (2003): 413–436.
26. "A Mechanically Scanned Deployable Antenna Using a Whiskbroom Feed System," Cont. NAS 5-26494, Harris Corporation, Melbourne, FL, February 1982.
27. R. R. Ferraro, F. Weng, N. C. Grody, L. Zhao, H. Meng, C. Kongoli, P. Pellegrino, S. Qiu, and C. Dean, "NOAA Operational Hydrological Products Derived From the Advanced Microwave Sounding Unit," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 5 (May 2005): 1036–1049.
28. J.R. Piepmeyer and A. J. Gasiewski, "High-Resolution Passive Microwave Polarimetric Mapping of Ocean Surface Wind Vector Fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 3 (March 2001): 606–622.
29. E. G. Njoku, J. M. Stacey, and F. T. Banath, "The SeaSat Scanning Multichannel Microwave Radiometer (SMMR): Instrument Description and Performance," *IEEE J. Oceanic Eng.*, vol. OE-5, no. 2 (April 1980): 100–115.
30. P. D. Potter, "A New Horn Antenna with Suppressed Sidelobes and Equal beamwidths," *Microwave J.*, (June 1963): 71–78.
31. M. C. Bailey, "The Development of an L-Band Radiometer Dual-Mode Horn," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (May 1975): 439–441.
32. R. E. Lawrie and L. Peters, Jr., "Modification of Horn Antennas for Low Sidelobe Levels," *IEEE Trans. Antennas Propagat.*, vol. AP-14 (September 1966): 605–610.
33. W. D. Burnside and C. W. Chuang, "An Aperture-Matched Horn Design," *IEEE Antennas Propagat. Symp.* (June 2–6, 1980): 231–234.
34. J. C. Mather, "Broad-Band Flared Horn with Low Sidelobes," *IEEE Trans. Antennas Propagat.*, vol. AP-29 (November 1981): 967–969.
35. J. Ruze, "Antenna Tolerance Theory—A Review," *IEEE Proc.*, vol. 54 (April 1966): 633–640.
36. T.-S. Chu and R. H. Turrin, "Depolarization Properties of Offset Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-21 (May 1973): 339–345.
37. T. Wilheit, *The Electronically Scanning Microwave Radiometer (ESMR) Experiment: The Nimbus 5 User's Guide* (Greenbelt, MD: NASA Goddard Space Flight Center, November 1972).
38. B. M. Kendall, "Passive Microwave Sensing of Coastal Area Waters," AIAA Conf. Sensor Syst. for 80s. Colorado Springs, CO, December 2–4, 1980.
39. Y. Rahmat-Samii, K. S. Kona, M. Manteghi, S. Yueh, W. J. Wilson, S. Dinardo, and D. Hunter, "A Novel Lightweight Dual-Frequency Sixteen-Element Stacked Patch Microstrip Array Antenna for Soil Moisture and Sea Surface Salinity Missions," *IEEE Antennas & Propagation Magazine*, vol. 48, no. 6 (December 2006): 33–46.
40. J. P. Hollinger, J. L. Pierce, and G. A. Poe, "SSMI Instrument Evaluation," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 5 (September 1990): 781–791.
41. T. Kawanishi, T. Sezai, Y. Ito, K. Imaoka, T. Takeshima, Y. Ishido, A. Shibata, M. Miura, H. Inahata, and R. W. Spencer, "The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASDA's Contribution to the EOS for Global Energy and Water Cycle Studies," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 2 (February 2003): 184–194.
42. F. J. Wentz, P. Ashcroft, and C. Gentemann, "Post-Launch Calibration of the TRMM Microwave Imager," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 2 (February 2001): 415–422.

43. P. W. Gaiser, K. M. St. Germain, E. M. Twarog, G. A. Poe, W. Purdy, D. Richardson, W. Grossman, W. L. Jones, D. Spencer, G. Golba, J. Cleveland, L. Choy, R. M. Bevilacqua, and P. S. Chang, "The WindSat Spaceborne Polarimetric Microwave Radiometer: Sensor Description and Early Orbit Performance," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 11 (November 2004): 2347–2361.
44. J. P. Hollinger, J. L. Pierce, and G. A. Poe, "SSMI Instrument Evaluation," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 5 (September 1990): 781–791.
45. C. S. Ruf, C. T. Swift, A. B. Tanner, and D. M. LeVine, "Interferometric Synthetic Aperture Microwave Radiometry for the Remote Sensing of the Earth," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, no. 5 (September 1988): 597–611.
46. M. Ryle and D. Vonberg, "Solar radiation on 175Mc/s," *Nature*, vol. 158 (September 1946): 339.
47. U. Bach, T. P. Krichbaum, E. Middelberg, M. Kadler, W. Alef, A. Witzel, and J. A. Zensus, "Spectral Properties of the Core and the VLBI-Jets of Cygnus A," *Proc. 7th Symp. European VLBI*, Toledo, Spain (October 12–15, 2004): 155–156.
48. S. J. Tingay, J. E. Reynolds, A. K. Tzioumis, D. L. Jauncey, J. E. J. Lovell, R. Dodson, M. E. Costa, P. M. McCulloch, P. G. Edwards, H. Hirabayashi, D. W. Murphy, R. A. Preston, B. G. Piner, G. D. Nicolson, J. F. H. Quick, H. Kobayashi, and K. M. Shibata, "VSOP Space VLBI and Geodetic VLBI Investigations of Southern Hemisphere Radio Sources," *Astrophys. J. Suppl. Series*, vol. 141 (August 2002): 311–335.
49. Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger, "Soil Moisture Retrieval from Space: The Soil Moisture and Ocean Salinity (SMOS) Mission," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 8 (August 2001): 1729–1735.
50. J. Font, G. S. E. Lagerloef, D. M. Le Vine, A. Camps, and O. Z. Zanifé, "The Determination of Surface Salinity With the European SMOS Space Mission," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 10 (October 2004): 2196–2205.
51. D. M. LeVine, T. T. Wilheit, R. E. Murphy, and C. T. Swift, "A Multifrequency Microwave Radiometer of the Future," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, no. 2 (March 1989): 193–199.
52. D. M. LeVine, M. Kao, A. B. Tanner, C. T. Swift, and A. Griffis, "Initial Results in the Development of Synthetic Aperture Microwave Radiometer," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 4 (July 1990): 614–619.
53. C. Ruf, C. Principe, T. Dod, B. Gosselin, B. Monosmith, S. Musko, S. Rogacki, A. Stewart, and Z. Zhaonan, "Lightweight Rainfall Radiometer STAR Aircraft Sensor," *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2 (June, 2002): 850–852.
54. J. R. Piepmeier and A. J. Gasiewski, "Digital Correlation Microwave Polarimetry: Analysis and Demonstration," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 11 (November 2001): 2392–2410.
55. B. Bizzarri, A. J. Gasiewski, and D. H. Staelin, "Observing Rain by Millimetre-Submillimetre Wave Sounding from Geostationary Orbit," *Measuring Precipitation from Space: EURAINSAT and the Future*, Chap. 50, series on *Advances in Global Change Research*, vol. 28, V. Levizzani, P. Bauer, and J. F. Turk (eds.) (New York: Springer, May 2007): 675–692.
56. F. Torres, A. B. Tanner, S. T. Brown, and B. H. Lambrigsten, "Robust Array Configuration for a Microwave Interferometric Radiometer: Application to the GeoSTAR Project," *IEEE Geosci. Remote Sensing Letters*, vol. 4, no. 1 (January 2007): 97–101.

Chapter 42

Antenna Tracking

Robert B. Dybdal

The Aerospace Corporation

CONTENTS

42.1	INTRODUCTION.....	42-2
42.2	TRACKING REQUIREMENTS.....	42-3
42.3	OPEN LOOP ANTENNA TRACKING	42-6
42.4	CLOSED LOOP ANTENNA TRACKING	42-10
42.5	ACQUISITION ISSUES.....	42-17
42.6	INTERFEROMETRIC SYSTEMS.....	42-19
42.7	ANTENNA TRACKING TESTING	42-22

42.1 INTRODUCTION

Antenna angle tracking systems have a variety of forms dependent on their application. System applications that use broad coverage antennas can have either no antenna tracking requirements or modest pointing requirements, such as the case in terrestrial relay designs. Other applications require very precise closed loop systems in order for high gain antennas to satisfy high angular accuracy requirements. Generally, antenna tracking systems are required when the uncertainty in the signal direction exceeds a small fraction of the antenna's beamwidth. This requirement derives from the need to maintain alignment of the antenna's peak gain level with the signal's direction to avoid signal loss. The boresight axis of the antenna is coincident with the antenna's peak gain level. At the outset, it is well to distinguish between *antenna pointing*, which is the capability to position the antenna's boresight axis towards a given direction (for example, an azimuth and elevation value), and *antenna tracking*, which is quite different. Antenna tracking is the capability to align the antenna's boresight axis (the one that intersects the peak gain level of the antenna's pattern) with the signal direction and then follow any variations in that direction. Thus, pointing is a static quantity whereas tracking is a dynamic quantity.

The system applications for antenna tracking include communication and radar systems. While both system applications share some commonality in techniques and implementations for antenna tracking, their requirements differ. Communication system signals generally have well-known directions of arrival for the signals and the received signal levels are generally well behaved with little variation in received power levels. By contrast, radar systems typically must detect potential targets, a phase generally referred to as *search and acquisition*, and then locate detected targets not only in the angular coordinates of interest using the angle tracking techniques discussed here, but also in the target's range and Doppler coordinates. Further, the received radar signal level typically has a large dynamic range, as the aspect angle between the radar and target changes.

A hierarchy of antenna tracking techniques exists. In some cases, the signal direction is stationary and antenna pointing rather than tracking is required. A common example is the user direct broadcast antennas that require only a simple pointing alignment. The simplest antenna tracking technique results when variations in the signal direction are known or predictable to an uncertainty that is a small fraction of the antenna's beamwidth. In this case, the antenna position can be simply commanded to follow known changes in the signal's directions, a technique known as *program track*. Program track is an example of an open loop tracking design, as is another technique referred to as *step track*, which when commanded, validates and updates correct antenna tracking. Closed loop antenna tracking techniques also exist, and two distinct techniques, conical scan, and monopulse systems, are described. These closed loop techniques continuously measure the signal's direction of arrival. The measured signal directions are an input to the antenna's control system, which commands the antenna to dynamically follow changes in the signal's direction. The closed loop techniques involve more than the antenna's RF design, and they include the means of pointing to and following the signal's directional variations using a positioner and control system. These features mechanically and/or electronically align the antenna's main beam with the signal direction in response to estimates of the signal direction derived from the antenna's response. As the antenna tracking system becomes more complex, test techniques to verify the performance of the antenna tracking design become more stringent, and such system testing is also discussed.

42.2 TRACKING REQUIREMENTS

The requirements for antenna tracking depend on the system's application. Communication system applications require maintaining signal alignment within a specified accuracy. Commonly, alignment within one-tenth of the antenna's beamwidth limits the signal loss to about 0.1 dB lower than the peak antenna gain level on the boresight axis. Radar systems have a more stringent requirement than signal loss and seek to determine the target's location with as much angular accuracy as possible. Generally, an angular accuracy of one-twentieth of the antenna's beamwidth or better is achieved by tracking radars. The selection of the antenna tracking design depends on the uncertainty of the signal direction relative to the antenna's boresight axis containing the peak gain level of the main beam. Commonly, the antenna's tracking accuracy normalized to the antenna's beamwidth¹ is the measure of performance. This tracking accuracy is the one sigma angular uncertainty normalized to the antenna's beamwidth. Error budget projections using the rss sum of the one sigma error component values and the algebraic sum of the mean components are used to quantify antenna tracking performance. It should be recognized that tracking accuracy does not have an absolute angular limit independent of the antenna's beamwidth. The tracking accuracy limit is a fractional part of the antenna's beamwidth.

Communication Systems

Antenna tracking is commonly required for larger ground terminal antennas used in satellite communication system applications. Satellite pointing directions are determined from the user location on the earth and the satellite's ephemeris, which describes the satellite's orbital motion. For geostationary satellites that maintain a fixed location above the earth's surface, the changes in the signal's direction is minimal and a slight figure eight motion over a day's time is encountered. The magnitude of this figure eight motion about the equatorial plane depends on how well the satellite is stationkept for maintaining its position on the equatorial plane and for compensating for orbital perturbations. The size of the figure eight is controlled by occasional satellite thruster adjustments. Satellites in other orbital configurations move in predictable trajectories described by the satellite's ephemeris and typically require positioners and antenna tracking systems to follow the dynamics of the satellite's trajectory.

The satellite communication application illustrates the variety of requirements for antenna tracking systems. Broad beamwidth antennas used with geostationary satellites require little more than antenna pointing alignment to limit signal loss, and a variety of such antennas for very small antenna terminals (VSAT) applications are commonly used. Positioning for these designs range from a fixed alignment to simple positioners that can be commanded to point the antenna to different satellites in the geostationary satellite belt. Antennas for other orbital configurations require more capable positioners that can follow the satellite's trajectory. The dynamic requirements of these positioners increase with reduced satellite altitude. Still more capable positioners are required for high data rate systems where the gain performance of a larger aperture is required for link closure. Such antennas can require open loop or closed loop antenna tracking design dependent on the antenna's beamwidth. Larger antennas are typically subject to pointing perturbations by wind unless they are protected by a radome. When unprotected by a radome, closed loop antenna tracking is generally required to compensate for wind disturbances.

The tracking accuracy requirement for communication systems limits signal loss resulting from antenna boresight misalignment. This loss depends on the antenna's main beam pattern and the sensitivity to misalignment relative to the antenna's boresight axis. Typical values

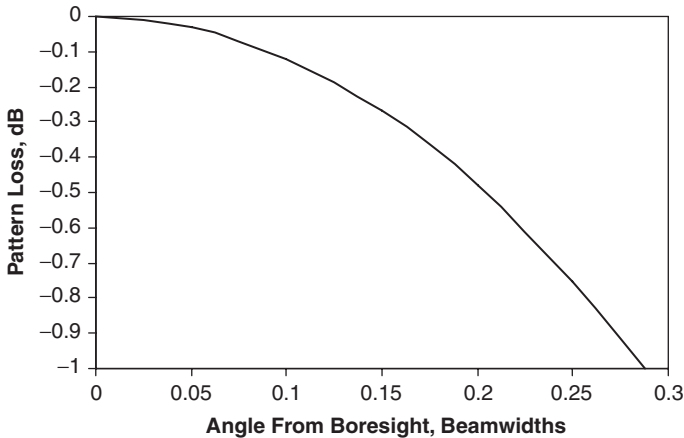


FIGURE 42-1 Pattern loss with boresight misalignment for communication systems

in Figure 42-1 illustrate the required tolerance of the signal loss to antenna misalignment, and they indicate a 1/10 beamwidth misalignment results in about a 0.1 dB signal loss. Antenna tracking is principally concerned with the antenna's pattern close to the main beam's boresight axis. A convenient means for determining this sensitivity assumes the shape of the antenna's high level portion of the main beam follows a Gaussian function at the higher gain levels. This is a good assumption for practical antennas that is used in this discussion. The pattern variation around the main beam peak, normalized to its peak gain level (voltage), is given by

$$F(\theta) = \exp(-K(\theta/\theta_{hp})^2)$$

where θ is the angular variation of the pattern in one plane, θ_{hp} is the half power beamwidth of the actual antenna, and $K = 1.3816$ as determined by evaluating the pattern at its half power point. This expression can be refined by using analytic projections of the actual antenna pattern predicted by commonly available computer codes or measured values of the actual antenna.

Larger antenna diameters that are not protected by radomes are susceptible to tracking disturbances caused by wind gusts. When the antenna aperture is buffeted by wind, the resulting torques disturb the antenna positioner system, producing antenna pointing errors. Thus, maintaining antenna pointing under wind loading depends in part on the torque capabilities of the positioner drives. Analysis techniques to determine antenna tracking errors caused by wind disturbances² have been developed and address the statistical disturbances of the antenna pointing as a function of the wind conditions. Ultimately, the wind torques are sufficient to potentially damage the antenna positioner and/or antenna, and vendor recommendations advise placing the antenna in a *stow* (pointing straight up); a position that minimizes the wind torque. Finally, vendors also specify a survival wind velocity in the stow position. When systems have high availability requirements, antenna radomes are typically used and the antenna is protected from wind-induced antenna pointing errors. The radome cost is offset in part by the reduced antenna structural and weather protection requirements and the torque requirements of the positioner drive.

Radar Systems

Radar systems³⁻⁴ can be broadly separated into *search radars* and *tracking radars*. Search radars look for potential targets and the antennas follow a fixed angular scan pattern. A typical example is airport surveillance radars that continuously scan both in azimuth and in range searching for aircraft. A time history of the radar returns defines the trajectory of the detected aircraft. Tracking in this case is really an open loop process. A plan position indicator (PPI) indicates target location in an azimuth direction and the radial distance from the indicator center is proportional to the target range. The time history of targets resulting from the persistence of this screen displays the trajectory of the aircraft target.

Search radar performance depends on the power aperture product of the radar, and therefore cost considerations favor low frequency operation. Tracking radars, on the other hand, gather more details regarding detected targets and the dimensions of tracking radar resolution include range, Doppler resolution, and, of interest in this discussion, target location in angular coordinates. Tracking radars generally operate at higher frequencies than search radars and use coherent radar waveform processing techniques to obtain improved resolution in comparison to search radars. The angular resolution of the target benefits from narrow antenna beamwidths, while Doppler resolution benefits from both the high frequency and coherent processing of narrow bandwidth waveforms. Finally, range resolution is achieved by processing wide bandwidth radar waveforms. In operation, narrow bandwidth and wide bandwidth radar waveforms are alternatively used to obtain range and Doppler resolution, respectively. In some cases, dual frequency radars are configured to provide search capability at the lower frequency and tracking capabilities at the higher frequency. A wide diversity of radar designs, with varying requirements for angle tracking, also exists.

Requirements for radar angle tracking differ from communication system requirements. Radar angle tracking strives to locate the angular coordinates of the detected target with as much precision as possible. Radar angle tracking requirements are more complex than antenna tracking requirements for communication systems, not only because radar systems generally require higher angular accuracy, more rapid acquisition times, greater dynamics in signal direction, and a dynamic range in received signal levels, but because the variation of the target aspect angle relative to the radar presents varying RCS (Radar Cross Section) levels from the target. This variation is received at signal level, and typically requires simultaneous signal reception of the main beam, azimuth, and elevation tracking beams commonly referred to as three-channel processing. In contrast to the one-tenth beamwidth accuracy typical for communication systems, the goal for radar angle tracking accuracy often exceeds one-twentieth of the antenna's beamwidth. Precision tracking radar systems typically use a large antenna at high frequencies to obtain narrow antenna beamwidths, high Doppler sensitivity, and sufficient waveform bandwidth for range resolution performance. They also use closed loop antenna tracking techniques to accurately measure the target location in angular coordinates.

Other issues for radar systems can arise. In some applications, multiple target tracking calls upon the radar to sequentially monitor the trajectories of several targets. Such monitoring requires rapid repositioning of the antenna beam. Antennas using reflector designs would require rapid mechanical motion to respond to multiple targets, and the time necessary to reposition and reacquire the targets may become excessive. In this case, the cost of array antenna technology may be justified to rapidly revisit targets with the array's electronic agility of beam positioning. The array technology has a limited scan volume over which the array's antenna gain can be maintained, and wider angular coverage typically requires separate array antennas covering different angular sectors so that radar system performance is not limited by array antenna beam scanning limitations.

Angle tracking at low elevation angles⁵ is problematic because of antenna multipath interaction and an image of the target being created by smooth surfaces such as water. Typically, at low elevation angles where the radar beam intercepts the terrain, the radar receives signals from both the actual target and its image. (The radar tends to track towards the horizon.) In such cases, operation at a sufficiently high frequency allows the terrain to be diffuse rather than specular. Higher frequency operation generally results in a narrow beamwidth antenna, and the minimum elevation position is kept above the terrain by one quarter of the antenna's beamwidth. Higher frequency operation also permits using sufficient waveform bandwidth to resolve the target from multipath and terrain images. These are effective techniques for maintaining angle tracking performance at low elevation angles.

Radar systems for terminal guidance applications have errors resulting from the target scattering properties. At long ranges, the accuracy of angle tracking is limited by S/N (Signal to Noise ratio). At close ranges to the target, scattered returns from different portions of the target can be resolved; typically, the radar is also in the near field of the radar target. As the radar approaches the target, the phasing between the different scattering centers comprising the target response varies, and this phasing results in different indicated target positions. This phenomenon is referred to as target "glint" and is the limiting error source for target tracking at close ranges. One treatment⁶ of this problem analyzes the balance point based on the target's scattering characteristics. This balance point is a centroid of the target tracking and dynamically moves as the individual scattering terms coherently combine in and out of phase; depending on the phasing, the radar alignment can be outside of the target dimensions. This treatment addresses the scattering mechanisms of the actual target by analytic means. A second approach⁷ examined the statistical behavior of measured tracking responses and proposes such statistical modeling in addressing glint issues. The statistics for a given target geometry, aspect angle, and radar frequency are required. Both of these approaches have instilled debate.⁸⁻⁹ Wide bandwidth radar waveforms can be effective in reducing this error source because the phasing effects between target scattering centers is decorrelated over the bandwidth.

In other applications, a tracking beacon is used on the target. The beacon has both receive and transmit antennas connected to a transponder, which amplifies the received radar signal. In contrast to skin tracking from the target's geometry, beacon tracking has some distinct advantages. Generally, amplification within the beacon provides a higher level return signal than the target itself reducing the performance requirements of the radar and/or increasing the radar's detection range. Additionally, the beacon return depends on the beacon's transmit antenna pattern, so that the radar's received signal has a better behaved variation with aspect angle changes than the skin return from the vehicle that generally has a lobe structure. Beacon operation is beneficial in satellite launch vehicle applications, for example, where tank separations and other events require the radar to determine the appropriate portion of the vehicle to track. In addition, the beacon also affords the opportunity to modulate other data onto the beacon return, providing trajectory, health, and other telemetry data. In some cases, the beacon translates the incident frequency both to provide additional isolation between the beacon's transmit and receive antennas and to separate the radar's receive signal from surrounding clutter returns in the normal radar response.

42.3 OPEN LOOP ANTENNA TRACKING

The most straightforward antenna tracking techniques are open loop designs. The open loop tracking technique follows the functional block diagram in Figure 42-2. The antenna is pointed using a positioner or in the case of an array steered to a nominal beam direction.

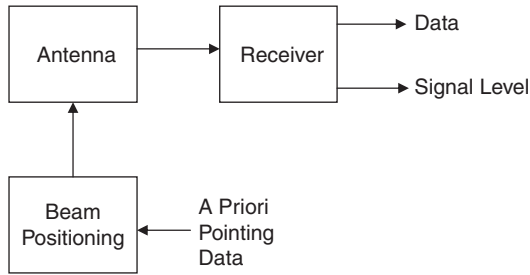


FIGURE 42-2 Functional diagram for open loop tracking

The receiver provides an indication of the received signal strength. The antenna beam is repositioned following *a priori* information. If the knowledge of the antenna's boresight axis positioning and the uncertainty of the pointing information corresponds to a small fraction of the antenna's beamwidth, such open loop methods can be quite effective and inexpensive to implement. The signal level can be the output of a carrier tracking loop or an AGC voltage. Two types of open loop tracking techniques are described: program track and step track.

The simplest form of antenna tracking is where the antenna's angular position is simply commanded to follow the time variation of the signal's direction, a technique referred to as program track. This technique relies on a known antenna location and orientation and the time history of the signal direction relative to the antenna's location. A common application is direct broadcast satellite services where the antenna is commanded to move from one satellite to another. The angular positions of the satellites defined by their ephemeris values and the antenna's location are known. This information provides the azimuth and elevation coordinates required to command the antenna to point to the selected satellite. Another application for relatively broad beamwidth antennas is meteorological satellites in a polar orbit. The satellite's trajectory is known and the antenna location is also known. This information allows the calculation of the time history of the required antenna positioning. The combination of the terminal's location, its orientation relative to level and true north, and time of day suffice to provide acceptable performance.

Step Track

While program track by itself is acceptable in many applications, another open loop tracking technique is useful in validating correct antenna tracking. This process is referred to as step track¹⁰ and involves pointing the antenna at its nominal position, commanding equal angular offsets in opposite directions, and then measuring the received signal power at these locations. If the nominal position is correct when the antenna is displaced in equal and opposite directions, the same signal power level is received. The operation of the step track technique is illustrated in Figure 42-3, where the nominal beam pointing is indicated, and the displaced beam pointings to both the left and right are also indicated. An example signal direction is indicated in the figure, and the pattern levels for the nominal and displaced beams are apparent. For this example, the signal direction is slightly removed from the boresight axis. If the antenna is moved in the vicinity of this point, small changes in signal level are observed and significant measurement accuracy is required to relocate the antenna to the peak gain value on axis. By contrast, the displaced beam positions result in significantly different signal levels.

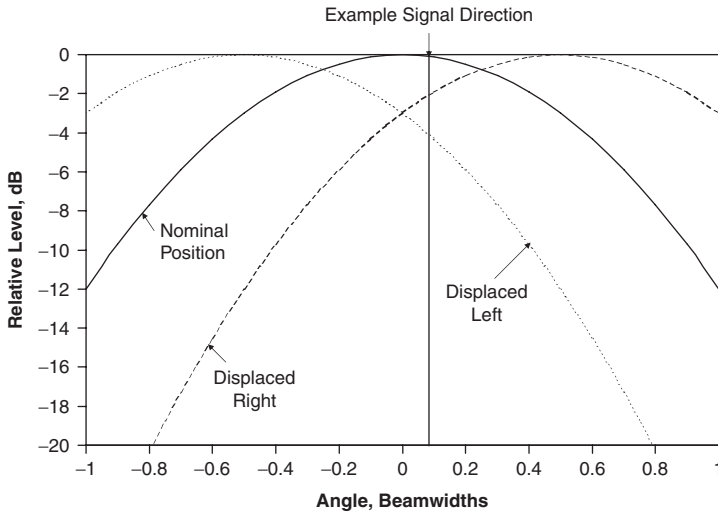


FIGURE 42-3 Step track operation

This difference in signal power levels allows the calculation of the angular correction needed to align the antenna with the signal direction. Assuming the Gaussian pattern representation, the angular correction θ_e can be easily computed from the amplitude imbalance¹⁰ at the two angular offsets using

$$\theta_e = -(\theta_{hp}^2/4K\theta_o) \ln R$$

where θ_{hp} is the antenna's half power beamwidth, θ_o is the angular offset from the antenna's boresight axis, K has been previously defined for the Gaussian function pattern approximation, and R is the amplitude ratio of the signal power at the two angular offsets. This process is repeated in the orthogonal plane.

In practice, step track validation of antenna tracking accuracy can be performed at regular intervals or by monitoring the received signal power (with an AGC output, for example), to determine whether or not an adequate power level is being received. For example, the antenna's pointing might be perturbed by wind gusts reducing the received signal power. Under this condition, step track can be used to decide if antenna realignment is appropriate. Another example is when a narrow beamwidth antenna tracks the stationkeeping variations of a geosynchronous satellite. Orbital perturbations result in geostationary satellites following a figure eight variation, centered on the equatorial plane over a 24-hour period. This movement is limited by correcting the satellite movement, a process referred to as stationkeeping. In this case, step track may be used to sample the figure eight movement at a discrete number of times over a 24-hour period. Because this motion repeats, the sampled points can be used to correct program tracking values.

In applying step track, the choice of the angular offsets, referred to as θ_o , must be made. This choice impacts the angular accuracy of the measurement. Small angular offsets have small reductions in signal power because the antenna pattern near the beam peak is relatively flat and high accuracy is required to measure signal level differences. Larger angular offsets experience more angular slope in the antenna pattern but the S/N (signal to noise ratio) is reduced which can increase measurement uncertainty. Thus, an optimum

value exists for the angular offset. A general expression¹⁰ for the step track accuracy σ_θ is given by

$$\sigma_\theta/\theta_{hp} = -(\theta_{hp}^2/2K\theta_o) \sigma_A$$

where σ_A is the one sigma accuracy with which the received signal amplitude can be measured. If the principal limitation of measuring the signal amplitude is S/N and that accuracy is $1/(2S/N)^{1/2}$, the accuracy becomes

$$\sigma_\theta/\theta_{hp} = k/(S/N)^{1/2}$$

where k is plotted in Figure 42-4. Notice that the tracking accuracy improves with increasing angular offset as a consequence of the increased slope of the main beam—until the reduced S/N limits the accuracy. An optimum value (minimum k) occurs at a pattern level of -4.3 dB. However, this is a broad optimum and in practical systems, different choices may be made. For example, most communication systems operate with a margin relative to threshold requirements as measured by bit error rate (BER) performance. In practice, the angular offset might be selected based on the signal margin requirements so that the data fidelity is maintained during tracking verification. It is important to recognize that the angular accuracy is limited by more factors than S/N considerations. For step track, positioner pointing errors and backlash, encoder quantization, and signal amplitude variations are other limitations. For this technique and others, it is important to construct an error budget incorporating all of the error sources specific to the applications to determine the overall accuracy of the design.

The step track antenna tracking technique as described has the assumption that the signal's direction remains fixed over the measurement interval. In some cases, this assumption is not valid, when, for example, tracking a polar satellite. Recently, the step track technique has been extended for this case and this extended technique is referred to as *rate corrected step track*.¹¹ One problem is the ability to sample the angular offsets while the signal direction changes in a known fashion, such as the case of satellite signals. In this case, the angular offsets are along the track and across the track of the trajectory rather

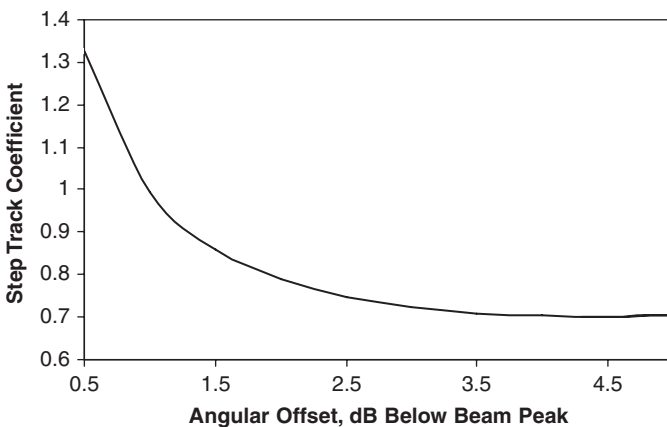


FIGURE 42-4 Step track accuracy coefficient versus angular offset (after R. B. Dybdal¹⁰ © IEEE 1998)

than repositioning the antenna in azimuth and elevation directions. An examination of the trajectory of a polar satellite reveals relatively small dynamics occurring at the rise and set of the satellite; the principal dynamics of the trajectory occur when the satellite is at the higher elevation angles of its trajectory. By using step track verification initially at the lower elevations, the ephemeris and antenna tracking can be corrected. The cross track motion centers the antenna on the trajectory, while the antenna beam location in the track direction is determined either by moving the antenna ahead and behind to identify the beam peak location or by moving the antenna ahead and slowing it to let the satellite drift through the main beam peak. As a practical matter, the satellite's orbital drag slows the satellite in orbit resulting in a time offset. Motion outside of the orbital plane requires perturbations other than drag. Thus, outdated ephemeris generally requires a time offset.

A second problem is the initial acquisition of the satellite as it clears the horizon; the multipath at low elevation angles limits tracking performance. The time required for initial acquisition is important because the goal is to collect as much data as possible during the time that the satellite is in view. To reduce the effects of multipath for step track operation, it is recommended that the antenna boresight be initially positioned above the horizon by about one quarter beamwidth, and azimuth motion be used to center the beam on its trajectory. As the satellite elevation angle increases, the satellite will drift through the main beam and the rate corrected step track can be initiated. By fixing the elevation angle, the phasing between the direct and multipath signal components tends to remain fixed.

These open loop tracking techniques are simple and cost-effective to implement. Such techniques are particularly useful for communication applications having a well-behaved signal power level. Their application to radar systems is limited by changes in the target's RCS levels during the signal sampling at the offset angular positions. Step track techniques are a useful means to verify the accuracy of program track in order to avoid excessive signal loss.

42.4 CLOSED LOOP ANTENNA TRACKING

Closed loop antenna tracking techniques require an antenna that has the means to dynamically measure the received signal's direction and a control system that can position the antenna based on the measured signal direction. The functional diagram in Figure 42-5 showing closed loop tracking differs from Figure 42-2's diagram for open loop tracking in several respects. The antenna must provide the capability for continuously measuring the

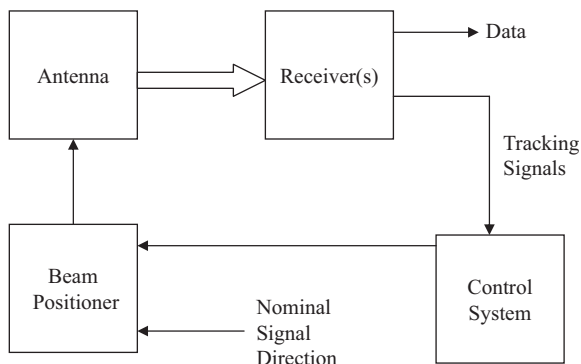


FIGURE 42-5 Functional diagram for closed loop tracking

received signal's direction in the form of a tracking signal. This directional information then provides an error signal that serves an input to the control system. By minimizing the error signal, the antenna is dynamically aligned with the signal direction. In operation, the antenna requires some *a priori* estimate of the signal directions and, for reasons that will be discussed, the uncertainty required in this *a priori* estimate is on the order of the antenna beamwidth. This *a priori* information can come from handover information from other systems or in the radar case, from target detection. Early closed loop tracking systems followed conical scan designs, while more recent designs favor monopulse designs. Closed loop antenna tracking techniques have been applied to both radar and communication applications.

Conical Scan

Conical scan designs¹ use a conventional antenna design and sense the signal direction by offsetting the boresight axis of the main antenna beam to coincide with a conical surface, and by rotating the antenna's boresight axis on the conical surface whose half angle corresponds to angular boresight offset. If the antenna is ideally aligned with the signal, the received signal strength is independent of the main antenna rotation on the conical surface. Antenna misalignment results in angular variation as the antenna's main beam traces its conical path. This angular variation has a sinusoidal form with its peak level occurring when the main beam is closest to the signal direction and a minimum level occurring when the main beam is farthest from the signal. A synchronous detector measures the received signal amplitude as the antenna beam traces its conical path. The amplitude variation increases with increasing antenna misalignment. The amplitude variation serves as an error signal for the control system, which operates by minimizing the amplitude variation. The magnitude of the amplitude variation increases with the magnitude of the misalignment. The azimuth angles of the maximum and minimum values (for example, the phase of the amplitude variation) indicate the direction to realign the antenna in order to reduce the amplitude modulation. In other words, the rotation axis of the conical scan should be repositioned in an azimuth direction from the highest value towards the lowest value.

Conical scan designs use a conventional antenna design and require mechanical motion to generate the conical scan pattern. For small antennas, such motion is commanded through the positioner drive. Larger antennas that have a dual reflector design generally offset and rotate the subreflector in order to generate the conical scan pattern with reduced mechanical rotation requirements, as compared to rotating the entire antenna. If conical scan is applied to array antennas, the beam rotation and repositioning is performed electronically. The angular accuracy for conical scan¹ is

$$\sigma_{\theta}/\theta_{hp} = 0.6/(S/N)^{1/2}$$

and is limited in part because the received signal is not aligned with the antenna's main beam peak gain level during tracking. Again, the angular accuracy is limited by error sources other than S/N , which must be addressed for specific applications. These error sources for conical scan include mechanical perturbations in the scan pattern and imperfect synchronization in the detection of the amplitude variation.

One problem with conical scan designs arises when the received signal amplitude fluctuates. This problem is particularly severe when the amplitude fluctuation rate is harmonically related to the conical scan frequency. Such variations in the received signal level result in a control system response that is unreliable. For communication systems requiring antenna tracking, the received signal level is generally well behaved. For radar systems, the received signal level depends on the target aspect angle relative to the radar, and changes in

the received signal level depend on the variation of the target's radar cross section (RCS) with aspect angle changes. This sensitivity to fluctuations in the received signal level limits the attractiveness of conical scan systems for radar applications.

An early variation of the conical scan design involves switching the antenna beam from one side of the boresight axis to the other side. This variation is referred to as sequential lobing, a technique that predates conical scan. Correct alignment results in the same amplitude at the switched positions, while misalignment results in different amplitudes. In early applications of sequential lobing, the control system was the operator. Sequential lobing has some features that are analogous to step track, but the lobing is performed continuously rather than being commanded at discrete time intervals like step track. Step track physically moves the antenna beam to off-axis position whereas sequential lobing electronically repositions the beams to off-axis positions. Like step track, sequential lobing is repeated in orthogonal planes. Like conical scanning, sequential lobing is susceptible to received signal amplitude fluctuations that limits tracking accuracy.

Monopulse Designs

Monopulse designs sense the signal direction by using two types of antenna patterns. One pattern is the conventional antenna response having a main beam aligned with the antenna's boresight axis and is used for data reception. The second pattern has a pattern null that is coincident with the antenna's boresight axis and is used for antenna tracking. The ratio of these two patterns, the tracking divided by the data pattern, yields a response that has a null on-axis and a linear slope (volts/volt/deg) for antenna misalignments from the boresight axis. This ratio is an error response and is the input to the control system for antenna beam positioning. The error response is the ratio of the tracking and data signals and results in the fundamental advantage of monopulse designs. Both the tracking and data signal levels can fluctuate with changes in the amplitude of the incident signal. However, their ratio is invariant with the incident signal fluctuations. Thus, monopulse designs are not limited by fluctuations in the received signal power. The term *monopulse* derives from the ability to obtain tracking information from a single radar pulse. True monopulse operation provides tracking information that is independent of signal fluctuations, unlike the sequential measurements made in other forms of antenna tracking.

Conceptually, antennas for monopulse operation¹² form the data and tracking patterns from two antenna beams symmetrically displaced from the antenna's boresight axis. When the two beams are added, an on-axis beam results. This beam is commonly referred to as the sum beam corresponding to the data pattern. When the two beams are subtracted, the pattern has a null on axis, and away from the axis, the beam is in phase with the sum beam in one direction and out of phase in the other direction. The two pattern responses using the Gaussian pattern assumption in Figure 42-6 illustrate typical pattern behavior. The ratio of the difference and sum beam is zero on axis, whereas near the axis it has a linear slope and is referred to as the *error response*, as illustrated in Figure 42-7. The error response follows the classic control system law in order to minimize the error that aligns the antenna with the difference pattern null. Since the peak of the data pattern is coincident with the error response null by design, minimizing the error response not only aligns the error response null with the signal direction, but also aligns the peak of the sum (data) pattern with the signal. Misalignment from the axis is measured by the error response magnitude and the sign of error indicates the direction towards the null.

The Gaussian pattern analysis can be used to derive the angular accuracy for monopulse systems.¹² This analysis assumes the two conceptual beams cross over at their half power point and it yields an angular accuracy given by

$$\sigma_{\theta}/\theta_{hp} = 0.51 \sigma_A$$

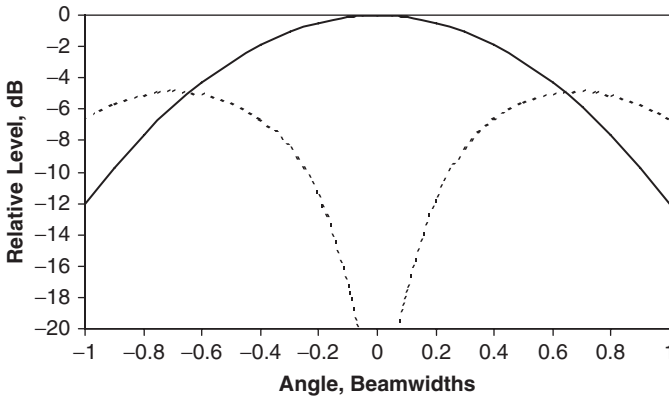


FIGURE 42-6 Monopulse sum and difference patterns

This analysis can be refined if the pattern data or error response of the actual antenna is available. Like step track, the angular accuracy depends on the ability to measure amplitude σ_A . Like step track, two limitations are S/N and positioning errors. Additionally, errors associated with the coincidence of the boresight axes of the sum and difference beams, the balance in the monopulse combining circuitry, and other factors need to be included in an error budget projection of tracking accuracy and also require quantification for specific designs. If S/N is the limiting factor in amplitude measurement accuracy, the monopulse accuracy becomes equal to

$$\sigma_{\theta}/\theta_{hp} = 0.361/(S/N)^{1/2}$$

Notice that the coefficient for the monopulse accuracy is smaller than those for conical scan or step track. Generally, monopulse tracking is more accurate than either conical scan or step track. However, the error budget limitations for the overall accuracy uncertainty differ

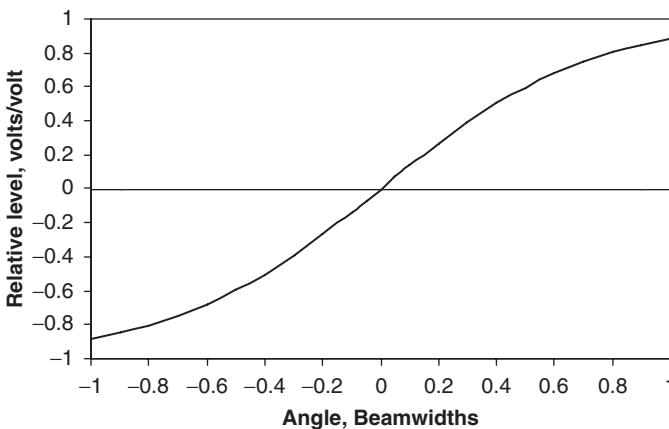


FIGURE 42-7 Monopulse error response corresponding to sum and difference patterns, as shown in Figure 42-6

in the three cases. In addition to the signal-to-noise issues, the monopulse errors include the balance of the feed circuitry, data and tracking pattern boresight coincidence, phasing errors, receiver tracking, and so on. This means that the individual error components in the uncertainty budgets must be addressed for all cases.

Monopulse designs require a more complex feed design than those used for open loop tracking, in order to produce the sum and difference beams. The problem of an optimum feed design has long been pursued. Early designs¹³⁻¹⁴ used multiple horns in the focal fields of the reflector to generate on-axis and off-axis beams, an approach that continues in radar designs. In radar applications, separate receivers are required for the sum, azimuth, and elevation difference beams so that measurements can be made simultaneously on both beam patterns. The multiple horn feed is also used in communication applications that does not have the same tracking sensitivity requirements as radar systems. Since communication systems generally have a well-behaved received signal level, the azimuth and elevation channels can be sequentially sampled and only a single tracking receiver is required. This approach is referred to as *pseudo-monopulse*. For communication applications, carrier recovery loops are often used to measure the incident signal amplitudes and the combination of a narrow bandwidth and averaging provide greater amplitude measurement accuracy than can be achieved from the signal alone. Accordingly, the difference channel antenna gain requirement can be lower than the sum channel requirement. One example for communication systems¹⁵ illustrates the integration of small feed apertures into the throat of a corrugated horn for tracking purposes.

Feeds using multimode excitations have developed more recently and are commonly used in communication applications. The communication application benefits from a well-behaved signal level in comparison to radar systems. Consequently, communication applications are often pseudomonopulse designs where the tracking signals are coupled onto the data signal and are sequentially sampled. This approach reduces the design complexity and the cost because separate tracking receivers and their amplitude balance requirements (each tracking receiver needs to produce the same output level for a given input signal level) are not needed. In practice, a dominant mode distribution in the feed provides the data or sum beam. A higher order waveguide mode sampled by a coupler supplies a tracking or difference pattern. The higher order mode has a null on axis as required for the tracking pattern. As the signal direction varies from the antenna boresight, the received signal level of the higher order mode increases, producing the linear variation from the null position required by the control system. The higher order waveguide modes, in addition to having an axial null, also have a phase progression with azimuth variations. When the dominant and higher order modes are combined, a peak level results when the two modes are in phase. The position of the peak level in azimuth is desired at quadrature points to provide the tracking response and then these quadrature points are sequentially sampled to determine the signal's azimuth position. If the amplitude is sampled at quadrature points using hybrid combining circuitry, azimuth and elevation difference channels are formed. In other cases, the magnitude and phase of the error response is measured. The magnitude indicates the magnitude to the boresight misalignment and the phase indicates the azimuth direction. In this case, the error response is a polar ρ and ϕ indication rather than an azimuth and elevation indication.

Three alternatives have been investigated to produce and process the dominant and higher order modes needed for monopulse operation. One approach¹⁶ uses multiple coupler arms to sample the higher order mode TE_{21} circular waveguide modes. The waveguide section containing the higher order waveguide coupler supports the dominant TE_{11} waveguide mode used for the sum signal. Eight TE_{21} mode couplers are spaced in an azimuth direction around the circumference of the waveguide. These eight arms are combined through hybrid circuitry to produce azimuth and elevation tracking outputs. Another approach¹⁷⁻¹⁸ uses the TM_{01} higher order circular waveguide mode and varies the azimuth angles where the

modes add in phase by using a phase shifter. This design uses four couplers spaced around the circumference of the multimode waveguide and then combines them with hybrids; an approach that has been demonstrated in wide bandwidth applications. A third approach¹⁹ uses a multiflare horn and a multimode coupler section and has a demonstrated high power transmitting capability. Orthomode junction couplers in the multimode section are used to sample the higher order tracking modes. The high transmit power capability in this case was obtained by combining four individual transmitters and the coupler design for the four transmitters was integrated into the feed design.

The tracking patterns obtained from the higher order waveguide modes and the data pattern obtained from the dominant mode must be coupled together in the pseudomonopulse design. The sequentially sampled tracking outputs provide an amplitude modulation on the data signal that is synchronously detected by the receiver. This detected amplitude modulation provides the input to the control system and the minimization of the amplitude modulation correctly aligns the antenna with the signal direction. One problem is that the path lengths through the tracking channels are generally longer than the straight path for the sum channel. Because of this, a "trombone" section is inserted into the sum path to equalize the path lengths between the data and tracking channels. This equalization of the group delay in the data and tracking channels is particularly important when polar ρ and ϕ processing is used. While a phase shifter can be used to equalize the phase, such an approach has a narrow bandwidth. When broad bandwidth antenna tracking is required, the phase shifter must be reset at discrete frequencies, if available for the broad bandwidth application.

The design issues for the multimode tracking techniques include the coupler that samples the higher order mode in a propagating section of the feed and the group delay differences between the sum and difference beams that result from their path length differences when these beams are combined together by a coupler. Couplers for the higher order modes are well demonstrated and the selection of the higher order mode to be used for tracking generally depends on the system's bandwidth requirement. This limitation results because the phase velocities of the dominant mode and higher order modes differ. The different group delay values of the sum and difference beams result in azimuth errors in the peak level of the dominant and higher order modes. These azimuth errors result in cross coupling degradation in tracking performance; for example, azimuth displacements produce indicated azimuth and elevation components. For narrow bandwidth applications, phase shifters can compensate for the group delay differences. For systems that operate at multiple frequencies, the phase shifter can be appropriately set for each of the bands. Such steps are not necessary if the waveguide lengths in the sum and difference channels can be equalized to produce the same group delay in both channels. By equalizing the group delays in the sum and difference channels, wide bandwidth operation can be achieved.

One problem with monopulse feed designs is their sensitivity to the polarization of the incident signal.²⁰ In polarization diverse signal environments, the feed response can lead to loss in tracking performance. For example, if the tracking channels of the feed are linearly polarized, the system will properly track circularly polarized signals of either sense with a polarization mismatch loss, but linearly polarized signals encounter difficulties when the polarization is cross-polarized to the design polarization of the feed. Because of the cross-polarization condition of the linearly polarized signal, the tracking signal does not measure a response and the system is without tracking information.

When the tracking channels of the feed are circularly polarized,²⁰ the system will properly track the design circular polarization with the proper design sense and all orientations of linear polarization with a polarization mismatch loss. However, the operation of the design when the system receives circular polarization of the wrong sense can actually become unstable. This instability results from the error response for cross-polarized signals. The error response for the design sense of circular polarization follows the standard response of having a null on axis and to first order a linear deviation as the antenna becomes misaligned

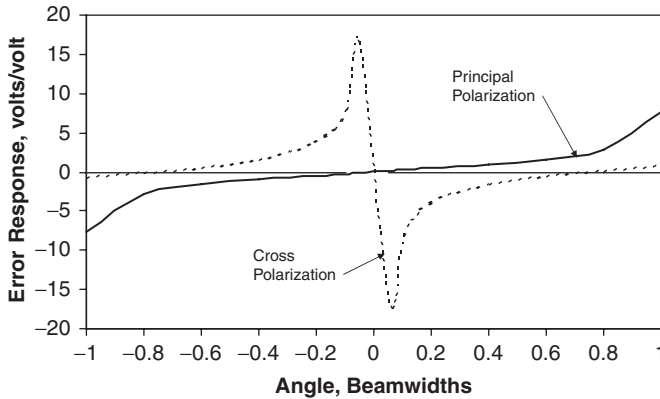


FIGURE 42-8 Error response for principal and cross polarization

from its boresight axis. Cross-polarized circular signals have an altogether different error response, having nulls in off axis positions corresponding to roughly the angular positions where the sum beam is roughly 10 dB below the beam peak. The error response differences between principal and cross polarizations are illustrated in Figure 42-8. The error response for the principal polarization case is well behaved and follows the desired linear behavior needed for the control system over a large fraction of the antenna's beamwidth. The error response for the cross polarized case has a much different behavior. Near the boresight axis, the error response peaks up and rapidly changes sign. This error response results in unstable tracking behavior as has been demonstrated.^{20,21} The dynamics of the antenna under such unstable operations are severe and the antenna tends to cone at an off axis position, probably where the error response nulls occur.

A technique²² has been devised to result in a tracking design that is independent of the signal's polarization for situations where the antenna must respond to a polarization diverse environment. The antenna is again configured for circularly polarized response for both senses of polarization. The tracking receiver is switched between both senses of circular polarization and the detected output for each polarization sense is summed to form an error response for the control system. If the received signal has the proper sense of the response being sampled, the proper error response is obtained. If the received signal has the wrong sense, the output is reduced in level by the polarization mismatch loss. The sum is dominated by the properly polarized response and the correct error response is routed to the control system. If the signal is linearly polarized, both the switched orthogonally polarized channels have the desired error response, albeit reduced by polarization mismatch. Such a design properly tracks both senses of circular polarization and linear polarization of any orientation.

A second problem is to make the sum channel independent of the signal's polarization, so that excessive polarization mismatch loss does not degrade signal reception. This is normally accomplished by combining the orthogonally polarized sum signal outputs with a diversity combiner. One configuration connects the two orthogonally polarized sum channels through a vector modulator, which combines the two channels in any arbitrary complex sum. Such diversity combining can be automated through an adaptive design. One approach examines the signal power differences for each polarization and combines the channels in both amplitude and phase in order to maximize the output signal power. A second approach uses a vector modulator that combines two signal paths in amplitude and phase. The modulator has two outputs, and by minimizing the signal power of one output through adjustment of amplitude and phase, the signal power is maximized at the second output.

Closed loop tracking dynamically follows directional changes in the received signal. These systems become necessary when the antenna's beamwidth is several times the uncertainty in the signal direction and accurate antenna pointing must be maintained. Such systems are more expensive than the open loop antenna tracking techniques. Part of the system expense lies with a more complex antenna feed design, because both data and tracking antenna patterns must be developed. Additional expense is incurred for the control system and the monitoring of the received signal levels. Radar systems must simultaneously sample the sum and two difference channels with three receivers to avoid degradation from received signal fluctuations. Communication systems generally have an easier and less costly implementation because the received signal level is well behaved. Thus, communication systems can sequentially sample a single receiver rather than multiple receivers and can effectively use averaging to maintain performance.

42.5 ACQUISITION ISSUES

Angle tracking techniques are successful when the received signal levels are sufficiently strong to allow detection. The process of aligning the antenna position with the signal is referred to as *acquisition* and assumes the system's receiver has an adequate signal level. Radar and communication systems again differ in the process of acquisition. The location of potential targets for radar applications is not known *a priori* and a search for the target must be undertaken. The location of potential targets for a tracking radar may be determined by handover information from a search radar or information regarding the target's trajectory may be known. Generally, tracking radars perform a limited search in the suspected target location, following a raster or spiral search pattern. The radar's range capability is a key requirement along with the target's radar cross section level. The radar search includes range and Doppler possibilities and an acquisition waveform may be used for coarsely determining the target's location. Radar target detection follows the usual issues of probabilities of detection and false alarm.

Communication systems are generally notified of the general signal direction and, as appropriate for the antenna's beamwidth, will conduct a limited search for the signal. Again, the antenna's beamwidth relative to the uncertainty of the signal direction dictates the need for a spatial search. Because the initial signal acquisition can often be done by acquiring the carrier component with the benefits of high sensitivity from a narrow bandwidth and averaging, the acquisition has greater sensitivity than the sensitivity needed to demodulate the data.

When extremely narrow beamwidth antennas are used, the spatial acquisition becomes more challenging. When very narrow antenna beamwidths are used, sometimes a separate, smaller acquisition antenna is used to provide a coarse determination of the signal's direction. While the gain of the smaller antenna is insufficient for data reception, the higher tracking sensitivity may still be able to obtain a coarse estimate of the signal direction to assist in acquiring the signal of the main antenna with its narrower beamwidth. A second approach²³ increases the acquisition field of view by using a cluster of antenna feeds in the focal region and individual tracking receivers on each feed. The spacing of these feeds is selected on the basis of the minimum required signal sensitivity, while a minimal number of additional feeds and tracking receivers can be selected to increase the acquisition field of view. The sensitivity of this approach is generally higher than using a separate, smaller acquisition antenna. Additionally, when one of the feeds acquires the signal, insight is provided to correctly align the antenna with the main beam from the knowledge of the pattern characteristics of the multiple beam of this design.

Antenna tracking techniques are limited to proper operation when the initial antenna positioning is within the main beam region. Away from this region, the anticipated tracking

response is not obtained. When narrow antenna beamwidths are used, the signal may be acquired with the signal arriving through the antenna's sidelobes. Three distinct techniques are available to verify the signal is aligned in the antenna's main beam rather than with an antenna sidelobe. One technique uses a separate smaller guard antenna. The guard antenna is selected so that this gain is comparable to the sidelobes near and adjacent to the antenna's main beam. A comparison of the signal level in the main and guard antennas are used to determine signal alignment. A single receiver can sequentially sample the signal level at the main and guard antennas. Separate receivers can simultaneously sample the received power levels, but the amplitude balance between separate tracking receivers, meaning both receivers have the same indicated output levels for the same input signal levels, must be maintained to produce reliable results. Signal injection through a coupler is one means to verify amplitude balance. If the main antenna signal level exceeds the guard antenna signal level, the signal is aligned with the main antenna's main beam and antenna tracking may be initiated. If the levels are comparable, a further search for main beam alignment is required.

A second technique²⁵ is based on the properties of the main antenna pattern and does not require a secondary guard antenna or tracking receiver. The angular width of the antenna's main beam is roughly twice the angular width of the antenna sidelobe. Measurement of the angular lobe width is used to distinguish main beam and sidelobe alignment. After the receiver has acquired, the antenna is commanded in angular offsets in order to measure the angular width of the lobe where the signal is received. This open loop commanding is similar to the operation of step tracking and if the location is determined to be the main beam, step tracking would be used for initial alignment. On the other hand, if this initial measurement indicates the smaller angular width of a sidelobe, the main beam is yet to be located. Open loop commands are then initiated to reposition the antenna to the next adjacent lobe position, and the angular width of this lobe is again measured. This process is repeated until the main beam is located. Insight into the pattern characteristics of the actual antenna is useful in applying this technique.

The third technique²⁵ applies to antennas having a monopulse capability. The error response has a linear response within the main beam and has high levels and a linear response having a different slope in the sidelobes, as shown in Figure 42-9. The error responses in this figure were determined by using $\sin X/X$ functions to obtain responses having a sidelobe structure. For the principal polarization, the error response follows the previously described behavior having a null on axis and a linear deviation away from the axis. This behavior persists over a significant portion of the main beam region and deviates from this response

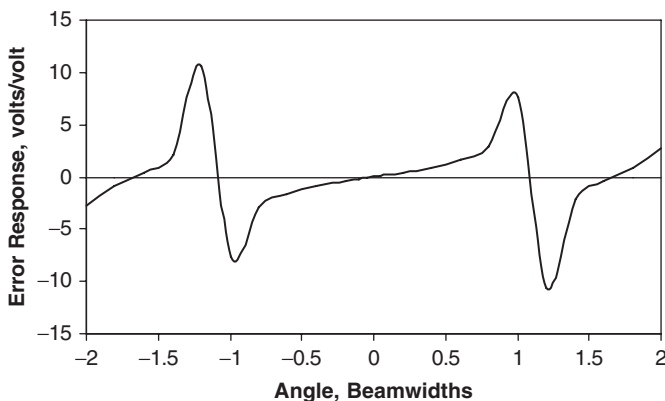


FIGURE 42-9 Error response over wider angular range

beyond it. In the sidelobe region, a portion of the response also has a linear slope, but the slope in the sidelobe region exceeds the slope in the main beam region. These differences in the error response can be used to distinguish alignment in the main beam region from alignment in the sidelobe region.

Generally, tracking receivers for the amplitude modulation containing the tracking response saturate above a predetermined percent modulation. If the tracking receiver has a saturated output, the signal is not aligned in its linear region and again commanded offsets can be used to realign the antenna to a point having a linear response. When a linear region of the error response is obtained, slight angular offsets can be used to determine if the slope of the error response is that of the main beam region or a sidelobe. If the slope is in the sidelobe region, angular offsets can again be used to search for the main beam. Once main beam alignment is verified, antenna tracking can be initiated.

42.6 INTERFEROMETRIC SYSTEMS

Interferometric antennas derive their tracking information by determining the phase difference between separated antennas. Such systems are sometimes referred to as phase monopulse, because the tracking information can be obtained from a single pulse and the relative phase between antenna elements is independent of the received signal level.

The phase difference between interferometric elements¹² for a signal arriving at an angle θ , with respect to the boresight direction of the elements, equals

$$\Delta\phi = 2\pi (d/\lambda) \sin \theta$$

where d is the baseline separation between the phase centers of interferometric elements and λ is the operating wavelength as illustrated in Figure 42-10. By taking the derivative

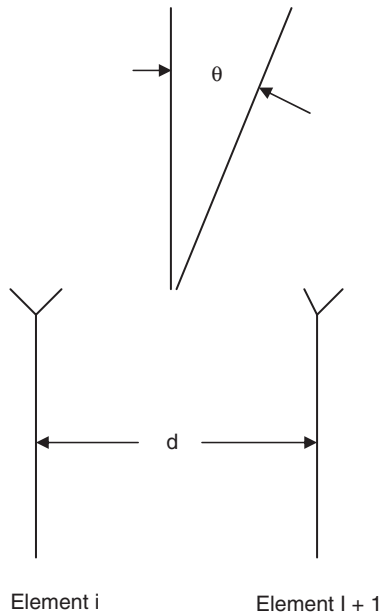


FIGURE 42-10 Geometry of interferometer

of this expression and noting that the half power beamwidth of the interferometric lobes equals

$$\theta_l = 0.88\lambda/(d \cos \theta)$$

it can be shown that the angular accuracy of an interferometric pair equals

$$\sigma_\theta/\theta_l = 0.181 \sigma_\phi$$

where σ_ϕ is the phase measurement accuracy. Thus, the angular accuracy of an interferometer is normalized to the beamwidth of the interferometric pair and depends on the accuracy with which the relative phase between interferometric elements can be measured.

In practice, the relative phase between antenna elements can be measured in a modulo 2π sense. Generally the received signal strength for each interferometric element dictates the required antenna gain level. Accordingly, the separation between antenna phase centers exceeds several wavelengths, because the antenna apertures cannot physically overlap. Because the phase between elements can only be in a modulo 2π sense, the signal direction becomes ambiguous, having several potential values when the baseline between the interferometric elements are several wavelengths. Interferometric ambiguities can be eliminated only if the baseline between the antenna element is less than one-half wavelength, and the angular accuracy for such a spacing is generally unattractive. Therefore, a means must be devised to reliably select the correct signal direction from the possible locations corresponding to the phase ambiguities.

One approach¹² to the phase ambiguity problem applies to interferometers that use high gain antenna elements. If these interferometric antenna elements are configured with a monopulse capability, the monopulse estimate can be used to resolve the ambiguities of the interferometer as illustrated in Figure 42-11. Thus, this system design enjoys

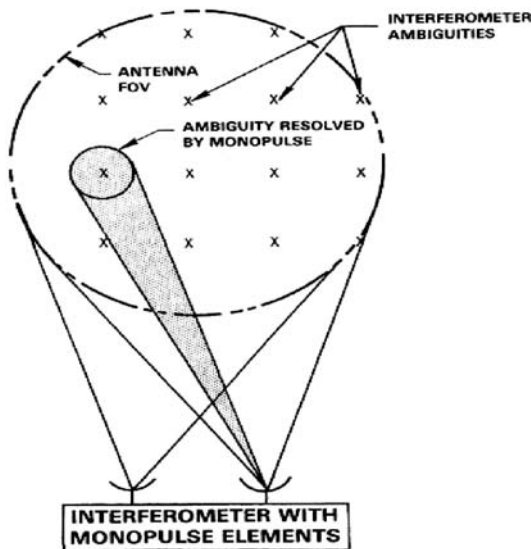


FIGURE 42-11 Monopulse resolution of interferometric ambiguities (R. B. Dybdal¹² © IEEE 1986)

the angular accuracy of the interferometer, while the monopulse capability provides the means to resolve the interferometric ambiguities. A probability of correct ambiguity resolution has been derived based on the diameters of the interferometric antenna elements, the baseline between the elements, and the accuracy of the monopulse capability. For practical applications, it can be shown that a high probability of correct ambiguity resolution exists if the element baseline is limited to four to five times the antenna diameter. An attractive feature of this approach is that the results are independent of frequency. As the frequency increases, the monopulse accuracy also increases maintaining the probability of correct ambiguity resolution. Thus, such a design can be used in wide bandwidth applications.

A second approach²⁶ applies in cases where low gain antenna elements are used in an interferometer. In this case, additional antenna elements are inserted within the interferometric baseline. In this approach, the overall angular accuracy is obtained from the longer baseline, and interferometric processing of the shorter baselines provides the means of resolving the phase ambiguities.

The strategy of selecting the spacing of the additional interferometric elements within the overall baseline must be devised. If the element spacing is an integer fraction of the total baseline, some of the ambiguities of the smaller baseline will be aligned with the ambiguities of the overall baseline. Those ambiguities that are not aligned can be eliminated as possible signal directions, and the process is continued until the correct signal direction is determined. The fractional integer value can be selected by determining the phase measurement accuracy of the interferometric pairs, and again a probability of correct ambiguity can be determined; practically, an element spacing of one-quarter to one-fifth of the total baseline or the next subbaseline as appropriate yields a reliable indication of the possible ambiguities. Like the monopulse resolution of phase ambiguities, this system can be used over wide bandwidths since the overall accuracy increases with increasing frequency as well as the resolution of the ambiguities. At the highest frequencies, the minimum element spacing may be insufficient to resolve all possible ambiguities and the physical constraints on the element spacing may not permit additional elements. In this case, forming a synthetic element by arraying two elements with the shortest baseline results in a phase center location between the two actual elements, and this baseline can be used to resolve the remaining ambiguity. A comparison of wide bandwidth techniques for angular location²⁷ illustrates the angular accuracy of interferometric approach and other monopulse techniques for broad coverage antennas.

Three distinct system parameters exist for antenna tracking designs: sensitivity or antenna gain, field of view, and angular accuracy. A further system constraint for practical systems is overall size. If the overall size is fixed, a comparison of monopulse and interferometric systems can be made. Monopulse design have a higher antenna gain because in a size constrained application, a single aperture is larger than multiple apertures required by interferometric systems; at the same time, the field of view for monopulse designs is accordingly smaller than interferometric designs. The angular accuracy of the interferometric design is better than the monopulse designs. Thus, the role for interferometry is applications where the wide field of view and angular accuracy requirements are more important than signal sensitivity. An example application for an interferometric design is a short range system where S/N is not a constraint and high dynamics in signal direction are anticipated. In this case, the smaller interferometric antenna elements have a wider angular field of view than a single monopulse antenna. This wider field of view is advantageous in following the angular position of the signal. Additionally, the reduced size of the interferometric antennas compared with the single large monopulse design eases the mechanical requirements for angular beam positioning.

42.7 ANTENNA TRACKING TESTING

Testing antenna tracking designs²¹ generally requires not only RF evaluation but also mechanical and control testing. The RF testing includes the normal evaluation of antenna gain, pattern, polarization, and bandwidth capabilities. The testing generally evaluates the feed performance and then the integration of the feed into the antenna. Monopulse testing is more complex, since the performance of both the sum and difference channels must be determined along with the error response characteristics. Careful attention to the coincidence of the sum channel's boresight axis and tracking channels' null is required. Attention must also be paid to the group delay differences between the sum and difference channels. Finally, the error response of the antenna, or the difference over sum pattern ratio, must be measured to establish the tracking sensitivity of the design and the angular region over which the linear error response prevails.

Having established the RF performance, the boresight axis of the antenna must be referenced to the mechanical axes of the antenna positioner. Small antennas can be measured in general purpose antenna test facilities and such testing does not necessarily require the operational positioner. The boresight alignment in this case is most conveniently performed in the test facility and the mechanical position of the antenna can be referenced to the electrical boresight axis, through the use of optical cubes, for example. Larger antennas, which cannot be readily accommodated in a test facility, require alignment when integrated with the positioner into the site. In addition to leveling the positioner, the positioner must be aligned to true north so that the angle encoders are properly registered.

Independent of how the antenna boresight is mechanically referenced to the positioner, any biases in antenna pointing must be determined. Signals whose directions are known are measured using the antenna system; such signals can include radio sources, signals of opportunity such as satellites, and calibration sources in known directions. The directions of these signals can be measured by step track and monopulse processing as appropriate. The antenna is commanded to point at the signal and differences in this direction and the direction indicated by antenna tracking provide the means to determine azimuth and elevation biases. Generally, antenna patterns are symmetric about the boresight axis near the main beam peak. Selecting different angular offsets (for example, 2, 2.5, 3, 3.5 dB) from the beam peak and looking at the boresight axis determined by the midpoint between the respective angular offsets provides some averaging opportunities for defining the boresight axis further. Least square fits to an approximate or measured beam shape is another alternative. The end objective is to verify the open loop pointing commands yield the same signal alignment as that determined by the antenna tracking technique. Independent signal of opportunities such as satellites can also be used to assure further confidence in the antenna boresight registration.

Monopulse systems alignment is more complex. The sum and difference patterns must have coincident boresight axes and open loop measurements are a starting point. The error response can also be measured and verification of null coincidence yields confidence. Measurement of the error response, and its deviation from linear, provides a measure of the acquisition range of the design. The need to have proper phasing between the sum and difference channels in order to minimize crosstalk between the azimuth and elevation channels also requires verification. An external signal source provides a tracking source; and signals of opportunity at relatively high elevation angles to avoid multipath should be used. (Such signals include existing satellite signals or radio sources.) The antenna is commanded to an offset angular position in azimuth, for example, and the tracking system is then activated. The azimuth and elevation encoders are observed and the tracking return to boresight should only occur in the azimuth direction in this example. Such tests are then repeated in the elevation plane. Testing of this type is commonly referred to as *snap on*

tests and should be repeated at different angular offsets to determine the angular limits of the tracking capability.

Other tests are required to evaluate the control system for closed loop tracking designs. The control system is evaluated by measuring its step response and ramp response.²¹ Again, a signal source is required for such testing and a low frequency function generator is used to inject offset error signals into the control system. The step response is measured by injecting a low frequency square wave into the error signal. The rise and fall characteristics define the control performance and the damping and overshoot characteristics are observed. The ramp response is obtained by using a triangular wave from the function generator and evaluates the linearity of the error response. By increasing the amplitude level of the function generator, the angular extent of the tracking system can be evaluated. This angular extent is limited by either the error response or when the amplitude modulation of the tracking signal saturates the tracking receiver. The dynamic range of the tracking receiver is measured by determining a threshold input level for acquisition and the input level needed to saturate the receiver.

The remaining testing requires a determination of the dynamic tracking performance. Such testing requires a moving signal source and knowledge of its dynamic motion. For satellite systems, the use of a polar satellite and measurements of the tracking during a satellite path provides such an opportunity. The satellite position is well known through the satellite's ephemeris and the time variation in azimuth and elevation can be determined a priori. Comparison of this time variation with the measured variation provides a dynamic evaluation of the performance. Selection of a satellite pass with a high elevation angle results in tracking trajectories having high dynamic motions and provides a good measure of the achieved satellite tracking performance.

The dynamics of the antenna are ultimately limited by the structural resonances of the design, which is a mechanical design issue. To first order from a control viewpoint, an antenna is an integrator until limited by structural resonance. A tabulation of roughly one hundred high performance antennas²⁸ has been made as a function of the antenna diameter. Examination of this data reveals the upper bound of the resonant frequency f_r shown in Figure 42-12.

$$f_r = 250/D, \text{ Hz}$$

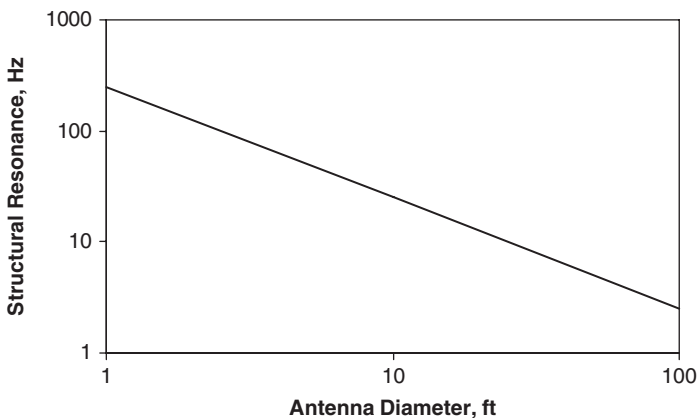


FIGURE 42-12 Structural resonance versus antenna diameter (after D. D. Pidhayny²⁸)

where D is the antenna diameter in feet. When evaluating the resonant frequencies, care should be taken to ensure other structural resonances associated with the antenna feed do not limit tracking accuracy.

Antenna tracking evaluation normally required the development of an error budget to project the uncertainty in antenna tracking accuracy. This budget is specific to the design and application of the antenna tracking design, the S/N obtained by the system design, and in the terminal guidance case, must also address the target effects. The error budget contains both mean and random error components and generally, the error budget is comprised of the algebraic errors and an rss sum of the random components. Such error budgets require significant attention to quantifying the individual error components. An example²⁹ illustrates the evaluation of a K-band ground terminal and includes the design description and evaluation including crosstalk evaluation and the effects of wind.

REFERENCES

1. D. K. Barton and H. R. Ward, *Handbook of Radar Measurement* (New Jersey: Prentice-Hall Englewood Cliffs, 1969).
2. ..., *Electrical and Mechanical Characteristics of Earth Station Antennas for Satellite Communications*, Electronic Industries Alliance/Telecommunication Industry Rept TIA/EIA-411 Revision A (September 1986).
3. E. Brookner, *Radar Technology* (Dedham, Massachusetts: Artech, 1977).
4. E. Brookner (ed.), *Aspects of Modern Radar* (Dedham Massachusetts: Artech, 1988).
5. D. K. Barton, "Low Angle Radar Tracking," *Proc IEEE*, vol. 62 (June 1974): 687–604.
6. L. Peters, Jr. and F. C. Weimer, "Tracking Radars for Complex Targets," *Proc IEE*, vol. 110 (December 1963): 2149–2162.
7. R. B. Muchmore, "Aircraft Scintillation Spectra," *IRE Trans Antennas and Propagation*, vol. AP-8 (March 1960): 201–212.
8. L. Peters and F. C. Weimer, "Concerning the Assumption of Random Distributions of Scatterers as a Model of an Aircraft for Tracking Radars," *IRE Trans Antennas and Propagation*, vol. AP-9 (January 1961): 110–111.
9. R. B. Muchmore, "Reply to Comments by Leon Peters, Jr and F. C. Weimer," *IRE Trans Antennas and Propagation*, vol. AP-9 (January 1961): 112–113.
10. R. B. Dybdal, "Step Tracking Performance and Issues," *1998 IEEE AP-S Symposium Digest* (June 1998): 58–61.
11. R. B. Dybdal and D. D. Pidhayny, "Rate Corrected Step-Track," *2002 IEEE AP-S Symposium Digest*, vol. 4 (June 2002): 752–755. See also R. B. Dybdal and D. D. Pidhayny, Method of Tracking a Signal From a Moving Signal Source, (May 4, 2004): U.S. Patent 6,731,240.
12. R. B. Dybdal, "Monopulse Resolution of Interferometric Ambiguities," *IEEE Trans Aerospace and Electronic Systems*, AES-22 (March 1986): 177–183.
13. P. W. Hannan, "Optimum Feeds for All Three Modes of a Monopulse Antenna," *IRE Trans Antennas and Propagation*, vol. AP-9 (September 1961): 444–461.
14. D. D. Howard, *Radar Handbook*, Chap. 18. M. I. Skolnik (ed.), (Boston: McGraw-Hill, 1990).
15. J. DiTullio, "A High Performance Feed for Earth Station Satellite Communications," 19th Annual USAF Antenna R&D Symposium Digest, Robert Allerton Park, IL, October 15–17, 1968.
16. Y. H. Choung, K. R. Goudey, and L. O. Bryans, "Theory and Design of a Ku-Band TE₂₁ Mode Coupler," *IEEE Trans Microwave Theory and Techniques*, vol. MTT-30 (November 1982): 1862–1866.
17. C. H. Chen, J. P. Pacente, and G. R. Simkins, Multiple Frequency Autotrack Feed for Wideband Communication Systems, (April 7, 1998): U.S. Patent 5,736, 907.

18. Y. H. Choung, "Wideband TM_{01} -mode Traveling Wave Coupler," *Proc IEE*, vol. 144 (October 1997).
19. K. R. Goudey and A. F. Sciambi, Jr., "High Power X-Band Monopulse Tracking Feed for the Lincoln Laboratory Long Range Imaging Radar," *IEEE Trans Microwave Theory and Techniques*, vol. MTT-26 (May 1978): 326–332.
20. R. B. Dybdal, "Polarization Limitations in Antenna Tracking," *2004 IEEE AP-S Symposium Digest*, Monterey, CA, June 20–25, 2004.
21. R. B. Dybdal and D. D. Pidhayny, "Evaluation of Antenna Tracking Systems," *2001 AMTA Symposium Digest* (October 2001): 139–143.
22. R. B. Dybdal and D. D. Pidhayny, "Polarization Optimization for Tracking Antennas," *2003 IEEE AP-S Symposium Digest*, vol. 2 (June 2003): 365–368.
23. R. B. Dybdal, "Extended Spatial Acquisition for Tracking Antennas," *1997 IEEE AP-S Symposium Digest* (July 1997): 2414–2417.
24. R. B. Dybdal, Extended Spatial Acquisition Method for Tracking Antennas, (September 14, 1999): U.S. Patent 5,952,962.
25. R. B. Dybdal and D. D. Pidhayny, "Main Beam Alignment Verification," *2004 IEEE AP-S Symposium Digest*, Monterey, CA, June 20–25, 2004. See also R. B. Dybdal and D. D. Pidhayny, Main Beam Alignment Verification for Tracking Antennas, (August 30, 2005): U.S. Patent 6,937,186.
26. R. B. Dybdal and P. R. Rousseau, "Resolution of Interferometric Ambiguities," *2000 IEEE AP-S Symposium Digest* (July 2000): 1964–1967. See also R. B. Dybdal and P. R. Rousseau, Method to Resolve Interferometric Ambiguities, (July 16, 2002): U.S. Patent 6,421,008.
27. L. G. Bullock, G. R. Oeh, and J. J. Sparagna, "An Analysis of Wideband Microwave Monopulse Direction Finding Techniques," *IEEE Trans Aerospace and Electronic Ssystems*, vol. AES-7 (January 1971): 188–202.
28. D. D. Pidhayny, The Aerospace Corporation, private communication.
29. T. Inoue and T. Kaitsuku, "K-Band Tracking Systems for Domestic Satellite Communication System," *IEEE Trans Aerospace and Electronic Systems*, vol. AES-17 (July 1981): 561–570.

Chapter 43

Microwave Beacon Antennas

Phillip N. Richardson

Texas Instruments, Inc.

Michael Carr

The Ohio State University ElectroScience Laboratory

CONTENTS

43.1	INTRODUCTION.....	43-2
43.2	INTERROGATOR ANTENNA DESIGN PRINCIPLES.....	43-3
43.3	INTERROGATOR ANTENNAS: PRACTICE	43-8

43.1 INTRODUCTION

A microwave beacon system, like a radar, transmits a pulsed RF wave to locate a target, using the time delay of the “echo” to determine distance and using antenna directionality to determine angular location. The distinguishing feature of a beacon system is that the target cooperates in this process, using on-board electronics to enhance the returned RF wave with amplification, frequency shifting, or coding. Beacons are thus highly accurate and reliable surveillance systems and also can provide some data-link capability. Beacon systems typically consist of transponders and interrogators. *Transponders* are the active devices carried by the target to provide the enhanced echo. Transponders are usually located on moving platforms, although fixed transponders may be used to mark hazards, navigation points, or calibration points. *Interrogators* employ equipment similar to that of conventional pulsed radars, i.e., a transmitter which elicits replies from the transponder and a receiver to detect and process the replies.

The most widely deployed beacon system is the military Identification Friend or Foe (IFF) system and its civilian derivative Air Traffic Control Radar Beacon System (ATCRBS). The military systems are also known as Mark X, Mark XII, and Mark IV IFF. Mark IV, the latest version, is known internationally as the NATO Identification System (NIS). The civilian systems are also known internationally as Secondary Surveillance Radar (SSR). All systems have similar waveforms and share common frequencies of 1030 MHz for interrogation and 1090 MHz for reply.¹⁻³ Polarization is always vertical. NIS also can be interrogated by certain radar signal formats at frequencies near C-band.

Transponders typically must respond to interrogations from any direction in space. Therefore, antennas with omnidirectional coverage in azimuth (see Figure 43-1) are desirable for this part of the beacon link. For aircraft, the antennas are usually blade antennas



FIGURE 43-1 Handheld Identification Friend or Foe (IFF) radar transponder with omnidirectional antenna (© 2007 Sierra Monolithics, Inc. All rights reserved.)

(aerodynamically shaped monopoles) protruding from the aircraft skin or flush-mounted annular slots in the skin. The design of these antennas is relatively straightforward, with primary emphasis on location to minimize shadowing.^{4,5} A modest 8 percent bandwidth covers the interrogation and reply frequencies. Interrogators, which also must respond to NIS radar interrogations, must have additional bandwidth. When aircraft shadowing prevents omnidirectional coverage from a single antenna, transponders are available with a time-sharing switch to rapidly sample two antennas (e.g., one on top of the aircraft and another on the bottom).

The interrogator may be either ground-based or airborne. In either case, the interrogator antenna is highly directional, with some means for scanning the directional beam over a volume of space. Thus, interrogator antennas are more complex in design and operation. These constitute the central focus of this chapter.

43.2 INTERROGATOR ANTENNA DESIGN PRINCIPLES

Requirements for interrogator antennas are similar to those for radar antennas, with the following differences:

- Beacons do not benefit from the “two-way” sidelobe suppression of radar.
- The beacon’s greater link reliability mandates greater attention to suppression of false targets due to multipath and sidelobe effects.
- Beacon power levels can be much lower because of target cooperation.
- Beacon antenna gain is often less critical than radar.

Minimization of false targets from multipath and sidelobe effects is a central design consideration. Collocation with a primary radar and the large physical dimensions imposed by the relatively low frequency are other factors.

Vertical Pattern Design

The preferred vertical pattern shape for surface-based interrogators is a sector pattern (uniform gain from the horizon to some prescribed upper limit) or a modified cosecant pattern. The cosecant pattern is preferred for long-range systems requiring high gain near the horizon. The upper limit of coverage is typically about 40° elevation.

Multipath reflection from surrounding terrain is a major contributor to false targets for surface-based interrogators. Multipath is minimized by shaping the vertical radiation pattern to reduce radiation below the horizon. Older antenna designs used small vertical apertures with broad vertical patterns, resulting in considerable radiation onto surrounding terrain. More recent designs have larger vertical apertures to shape the vertical pattern for substantially reduced radiation below the horizon. Figure 43-2 shows examples of vertical lobing for small and large apertures.

The antenna beam is usually tilted up slightly so that the pattern level at the horizon is –6 dB. A useful measure for multipath rejection is the slope of the vertical pattern at the horizon, measured in decibels per degree. Figure 43-3 shows the envelope of the multipath minima as a function of this cutoff rate. Since higher cutoff rates require larger apertures, the trade between antenna size and reduced multipath must be made carefully. A cutoff of 2 to 4 dB per degree has been found to be satisfactory for most installations.

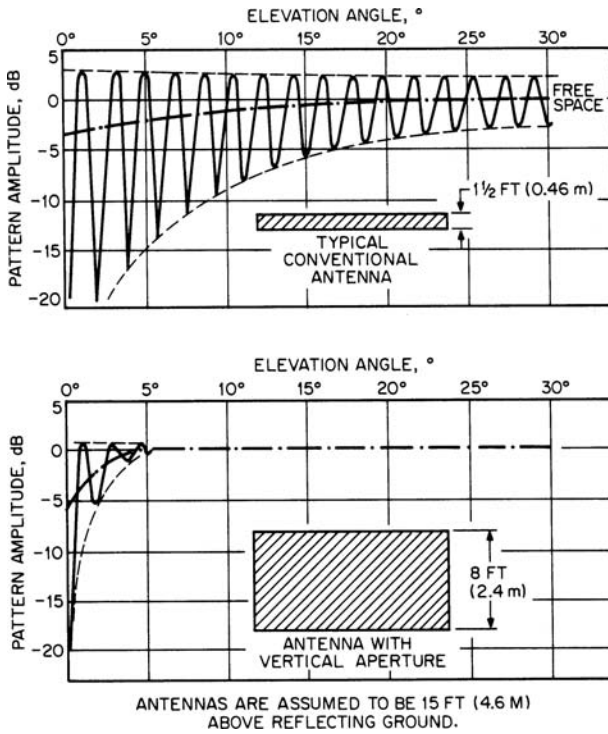


FIGURE 43-2 Influence of vertical aperture on elevation lobing patterns

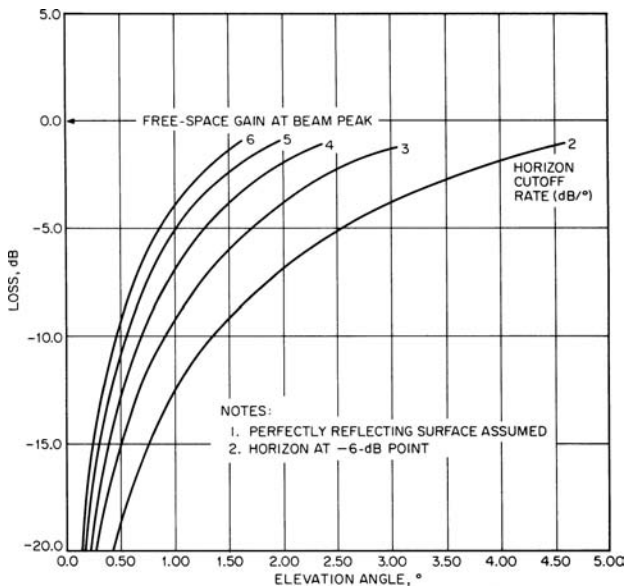


FIGURE 43-3 Envelope of lobing minima for various horizon cutoff rates

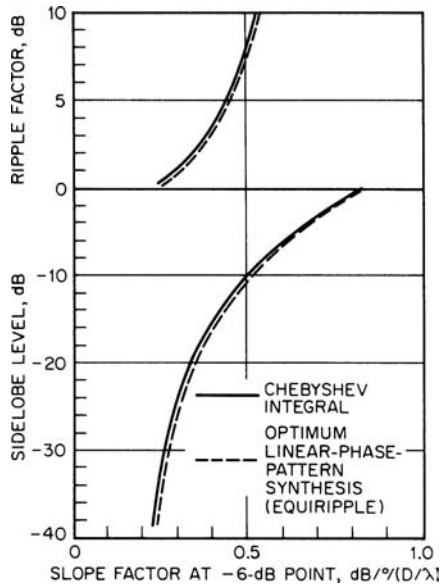


FIGURE 43-4 Normalized horizon cutoff rate (after Lopez⁷ © IEEE 1979)

Pattern-synthesis techniques have been developed to optimize horizon cutoff subject to constraints on antenna size, above-horizon coverage, and sidelobe levels. For the synthesis of sector beams, Evans⁶ has adapted a procedure originally developed for digital filters. For engineering estimates, Lopez⁷ provides curves (normalized to a vertical aperture of one wavelength) relating horizon cutoff, sidelobe level, and ripple within the sector beam (see Figure 43-4).

For urban interrogator locations, limited electronic scan of the vertical pattern is very useful to avoid direct illumination of tall buildings, thereby reducing reflected false targets. This feature, called *beam hopover*, allows the -6-dB pattern point to be tilted up slightly when interrogating in the direction of the obstruction. Array implementations of the interrogator can use a multiple-beam feed network to achieve this feature with a minimum number of switches or phase shifters. With a reflector antenna, a secondary feed below the focus can achieve the result.

Recent designs have used dipole arrays because they provide the control of vertical illumination needed to obtain shaped patterns, good horizon cutoff, and effective use of vertical aperture. Shaped reflectors also can be used, with multiple-horn and line feeds.

Because the vertical pattern is broad, it provides no significant ability for vertical-plane direction finding. This is usually not a limitation, because many transponder replies include a coded report of altitude. Thus the interrogator often receives a direct report of the target's vertical coordinate.

Horizontal Pattern Design

In the horizontal plane, accurate determination of target angle and minimization of false targets are of central importance. Surface-based interrogators typically use a low-sidelobe beam with a 3-dB width of 2 to 4°. The target's location can be measured to an accuracy of about one-tenth the 3-dB beamwidth by scanning the beam across the target

while interrogating continuously, with the target location estimated as the centroid of the “run” of replies that is received. For greater accuracy, a monopulse difference pattern is included in the horizontal plane. Standard off-boresight monopulse processing is used. With monopulse, angle accuracy of 1 to 2 percent of the half-power beamwidth is possible.⁸ For monopulse antennas, the two patterns are called the Σ (sum) and Δ (difference) beams, respectively.

For arrays, the Σ beam is formed with cophased illumination in the horizontal dimension. Amplitude tapering for -25 - to -30 -dB sidelobes is typical. The Δ beam is obtained by feeding opposing sides of the aperture in phase opposition. Use of the Σ amplitude taper for both the Σ and Δ patterns gives a simple feed network but results in quite high Δ sidelobes. This high-sidelobe Δ pattern is suitable only for noncritical uses. A more complex multibeam feed network, capable of producing an independently optimized Δ illumination (such as the Bayless taper), is usually required. Optimal Σ and Δ illuminations are very similar on the outer parts of the array, so the multibeam feed is only required for the center portion of the array. Figure 43-5 shows the two types of difference patterns and their respective aperture illuminations.

For reflectors, a conventional parabolic contour in the horizontal plane gives the desired pencil beam. Careful control of feed spillover is needed to ensure that backlobes are -30 dB or lower. Standard monopulse feed design will give suitably low Δ sidelobes.

Sidelobe Suppression

A procedure called *interrogation sidelobe suppression (SLS)* cues the transponders to avoid replying to sidelobes of the interrogator antenna. With SLS, the interrogation is transmitted first by the Σ pattern and then by an omnidirectional “control” pattern (Ω). The Ω gain is lower than that of the Σ beam but higher than that of the sidelobes. The transponder replies only if the amplitude for the Σ interrogation exceeds that for the Ω by some prescribed margin, indicating that the interrogation came from the main beam. A variation called *improved*

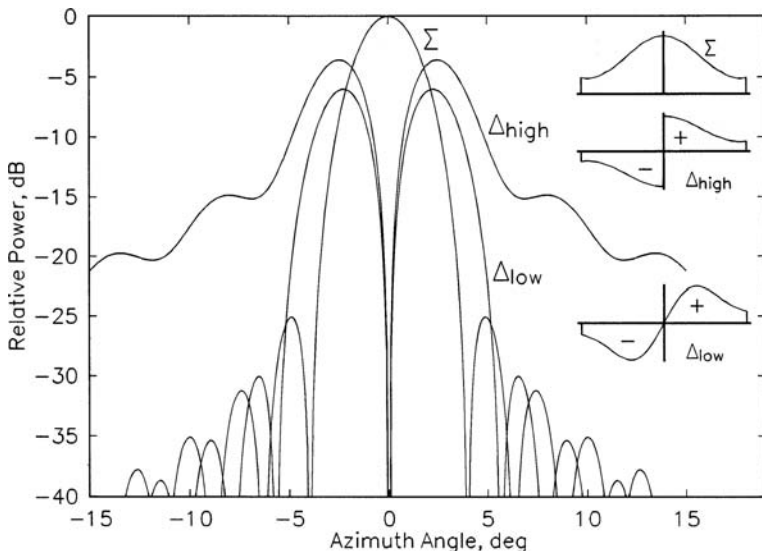


FIGURE 43-5 Typical far-field horizontal patterns with the corresponding aperture distributions shown in the insets

SLS (ISLS) provides for suppression of reflected main-beam interrogations. With ISLS, a pulse is transmitted simultaneously on the Σ and Ω beams, and the transponder measures timing differences to detect an interrogation arriving via a reflected path. Some beacon systems also employ SLS on the reply link (RSLS). With RSLS, a two-channel receiver compares replies received on the Σ and Ω antennas, blanking the reply when the relative amplitude indicates a pulse received by a Σ sidelobe.

When the directional and control beams (Σ and Ω) are derived from the same aperture, they share a common phase center. Consequently, the multipath lobing (see Figure 43-1) will be the same for both beams, so the proper Σ/Ω amplitude is preserved at all elevations. If, however, the phase centers are vertically displaced, the lobing does not match, and faulty SLS operation occurs. When an Σ peak aligns with an Ω null, sidelobe amplitude may be high enough to “punch through” the Ω coverage. When an Σ null aligns with an Ω peak, even the main beam may be unable to exceed the Ω gain, and all replies are suppressed. This differential lobing can be a serious problem, so most recent designs have used common phase centers.

Common types of Σ and Ω patterns are illustrated in Figure 43-6 and described below.

- **Omnidirectional pattern** This pattern is usually formed by a separate antenna, often a collinear array of dipoles. Its advantage is simplicity. It does not rotate and therefore does not require an additional rotary joint channel, and it does not constrain the design of the

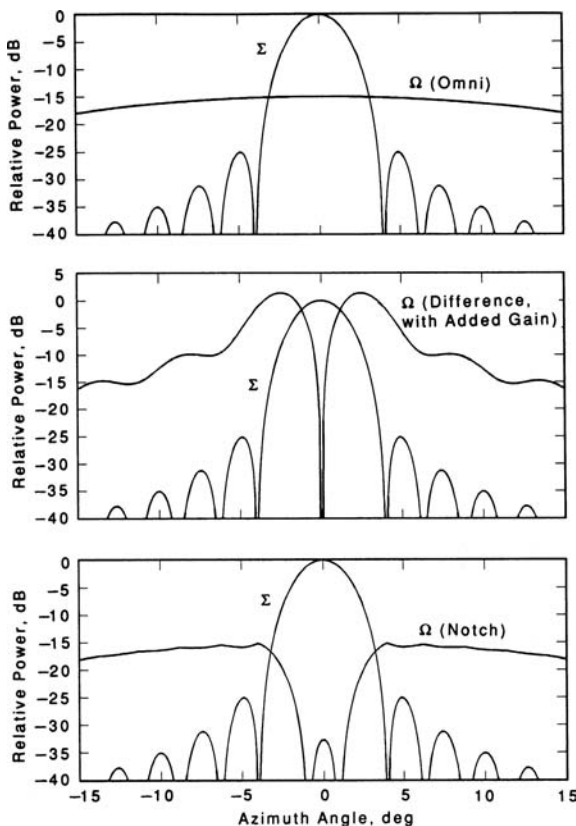


FIGURE 43-6 Generic control-pattern types

directional antenna. The vertical pattern of this Ω antenna is often designed to match that of the Σ . To avoid mutual blockage, the two antennas are displaced vertically, causing differential lobing. Such systems exhibit marginal SLS operation^{9,10} and are avoided in critical applications.

- **Difference pattern** The high-sidelobe difference pattern formed by an abrupt phase reversal at the center of the Σ illumination is sometimes used for Ω coverage. The vertical patterns and phase-center locations are closely matched, so differential lobing is minimized. The pattern is easily formed, requiring only a “magic T” power divider at the center of the array, and it imposes no other design constraints. Its principal shortcoming is that it provides little Ω coverage behind the antenna, allowing possible interrogation by backlobes and reflections.

With this pattern, the Ω pulse can be transmitted at a higher power level relative to the Σ because the null precludes main-beam blanking. The higher power compensates for the relatively poorer Ω coverage. This also narrows the effective beamwidth over which replies are elicited, because a substantial part of the main beam skirts are also blanked. Airborne military interrogators, where aperture size is often limited, use this effect to improve angular accuracy. For example, with Ω power that exceeds Σ by 5 to 10 dB, an antenna with a 35° 3-dB beamwidth can elicit replies over only a 5° sector. The design must consider the substantially higher power at the Ω port.

- **Notch pattern** For array implementations, this type offers an excellent combination of common phase center, central null, and control of sidelobe punch-through. The pattern is produced by the broad radiation from a single central column of the array, with simultaneous out-of-phase by the Σ pattern to produce the notch. Because of the Σ 's greater gain, little power is needed to produce the notch, typically about 1 percent of the total. An auxiliary “backfill” antenna is usually added to cover backlobe radiation in this approach. (A backfill antenna is also useful with the difference-pattern omni.)

43.3 INTERROGATOR ANTENNAS: PRACTICE

The following paragraphs describe practical issues faced by designers of ground- and air-based beacon interrogation antennas.

Ground-Based Systems

For many years the standard configuration for ATCRBS/IFF interrogators consisted of a linear array (commonly called a *hog trough* because it resembles an animal feeder) and a stationary omnidirectional antenna. The hog trough is a sectoral horn fed by a horizontal array of dipoles and is about 26 ft long. The beamwidth is 2.3° in azimuth and 42° in elevation. Peak sidelobe level is typically -23 dB. The omnidirectional antenna is a biconical horn or collinear dipole array mounted on a fixed mast beside the interrogator. Figure 43-7 shows such an installation at a Federal Aviation Administration (FAA) site. A more recent design,¹¹ shown in Figure 43-8, features an integral notch-type SLS pattern and also a difference pattern that can be used for monopulse angle estimation and RSLs. The military uses various shorter versions of the hog trough for its mobile interrogators. An example is shown in Figure 43-9. These are approximately 14 ft long, have a difference-pattern Ω , and have a wider azimuth beamwidth, typically about 4°. All of these antennas have little elevation beam shaping and little horizon cutoff (about 0.3 dB per degree). All are subject to problems of multipath.



FIGURE 43-7 Typical hog-trough antenna (*Courtesy of Department of Transportation, Federal Aviation Administration*)

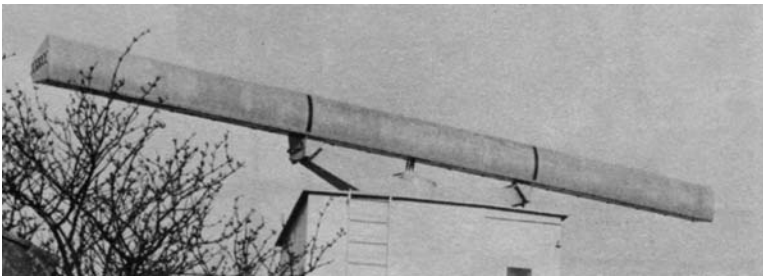


FIGURE 43-8 Typical SSR antenna (*Courtesy of Cossor Electronics, Ltd.*)

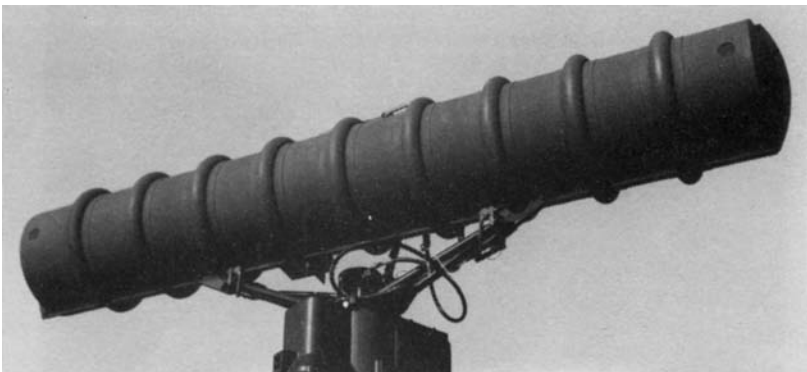


FIGURE 43-9 Typical military IFF antenna (*Courtesy of Hazeltine Corporation*)



FIGURE 43-10 Open-array SSR antenna. The ladder-like top section makes up the SSR directional antenna.

More recently, designs have employed increased vertical aperture to control multipath. The *open-array antenna*¹² shown in Figure 43-10 is a planar array of dipoles with a vertical aperture of 5 ft. It has a horizon cutoff of 2.2 dB per degree and employs an integral notch-type Ω pattern (with backfill radiator) for SLS, so multipath problems are greatly reduced compared with the hog trough. The dipole array is made unidirectional by columns of parasitic reflector dipoles located behind and between the columns of driven dipoles. This eliminates the need for a ground plane, giving a very open construction that has low wind loads, even with ice accumulations. Typical radiation patterns for this antenna are shown in Figure 43-11 and Figure 43-12. The open array has replaced the hog trough as the standard ATCRBS antenna at FAA terminal radar sites.¹³

At sites which have a large radar reflector as part of the surveillance equipment, a sharp horizon cutoff for the beacon can be obtained by sharing the reflector with the radar, using an *integral beacon feed*. Since both feeds must be located approximately at the reflector's focal point, the principal design problem is ensuring that each feed operates properly in

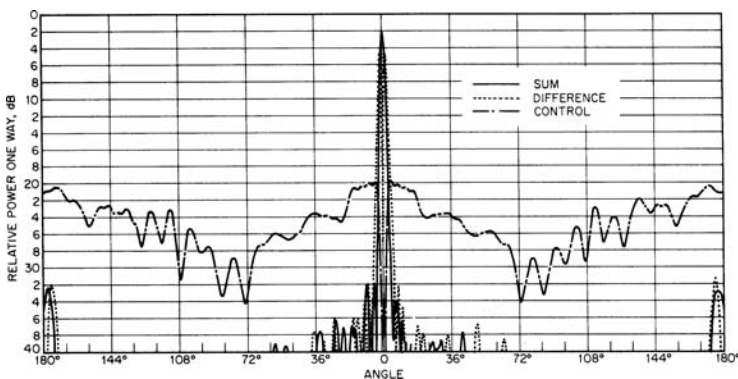


FIGURE 43-11 Open-array azimuth patterns (Courtesy of Texas Instruments, Inc.)

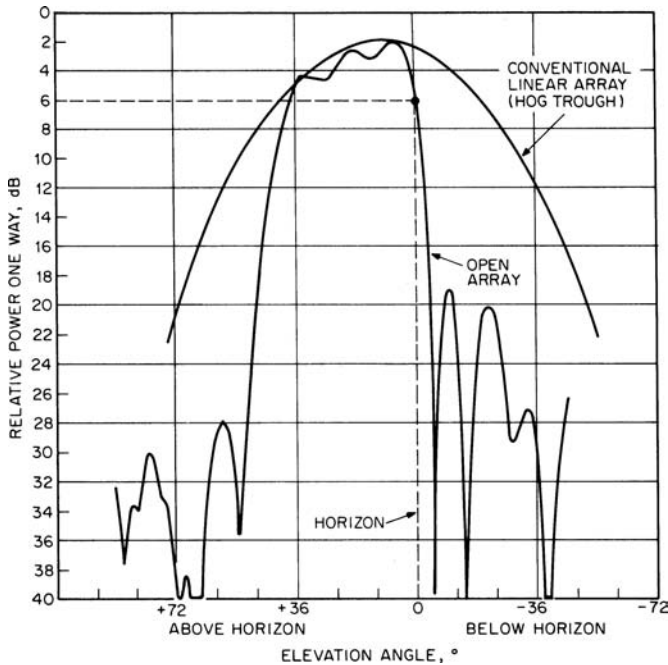


FIGURE 43-12 Open-array elevation pattern (Courtesy of Texas Instruments, Inc.)

the presence of the other. S-band and L-band radars are commonly associated with beacon systems. The two bands have somewhat different design concerns, as follows:

- The S-band reflectors are relatively small, yielding approximately a 4° azimuth beamwidth and 2 dB per degree horizon cutoff. The 4° beam is too wide for most FAA applications but is useful for certain less-demanding military applications. The S-band feed horn is small relative to the longer beacon wavelength, so an array of beacon dipoles can usually be disposed around the horn to make a beacon feed whose phase center is at the focal point.
- The L-band reflectors are larger, giving a 2° azimuth beamwidth and 4 dB per degree cutoff. However, the larger physical size of the L-band feed horn usually precludes a separate dipole array around the radar horn. Therefore, either the beacon must be duplexed into the radar horn or the beacon feed must be displaced from the focal point.

For either band, the beacon feed must be isolated from the radar so that it is not damaged by high radar-power levels and so that radar power is not coupled into the beacon receiver. In many cases the radar feed is circularly polarized, with both amplitude and phase patterns carefully adjusted for good axial ratio and proper illumination of the reflector. The beacon feed, being vertically polarized, can alter the axial ratio, degrading polarization performance of the radar. For either band, the elevation pattern is largely determined by the reflector contour, and it often approximates a cosecant-squared shape. The beacon feed pattern can be modified to direct more energy at the “spoiled” part of the parabolic contour, thereby increasing elevation coverage at higher angles to more nearly approximate a sector pattern.

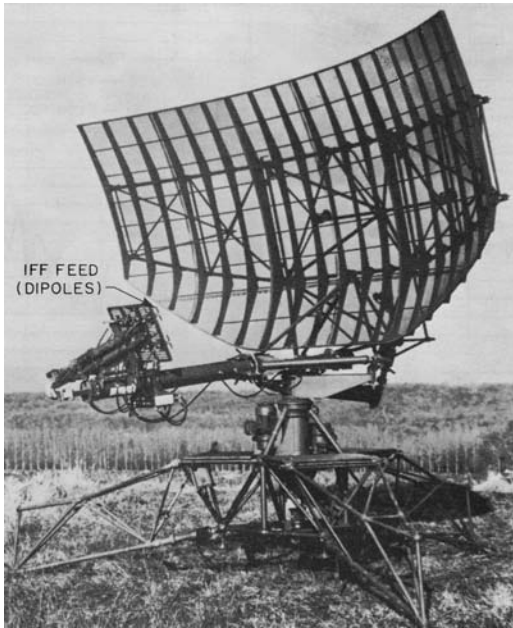


FIGURE 43-13 Tactical S-band radar antenna with integral IFF feed (*Courtesy of Thomson-CSF*)

Figure 43-13 shows beacon dipoles around the feed of an S-band radar. An array of dipoles, as shown in Figure 43-14, provides the feed-pattern control needed to optimize the beacon's azimuth and elevation-pattern shapes.

In the L-band ARSR-3 radar, the beacon feed is located beside the radar horn, off the focus (see Figure 43-15 and Figure 43-16). This introduces an azimuth offset between the

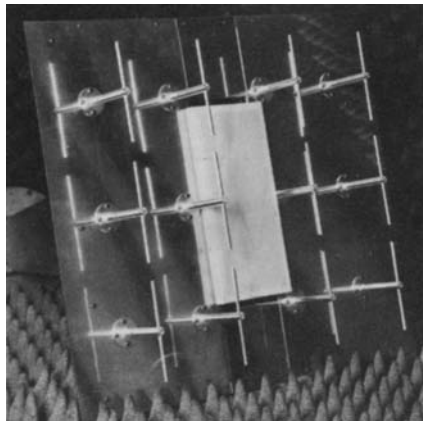


FIGURE 43-14 Integral beacon feed for elevation-pattern shaping (*Courtesy of AEG-Telefunken*)

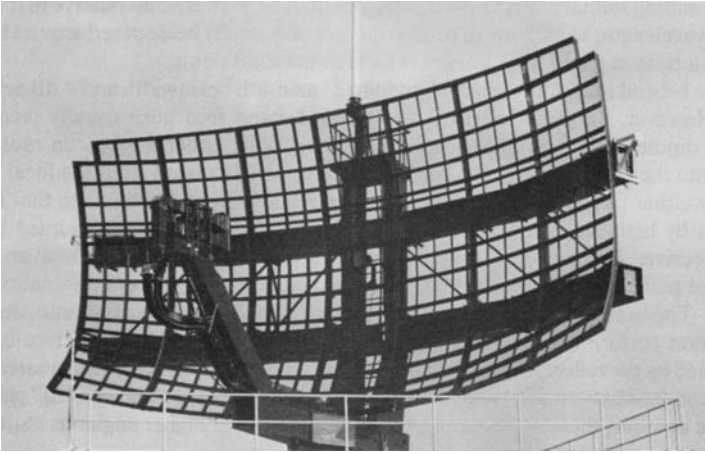


FIGURE 43-15 ARSR-3 long-range surveillance radar antenna (*Courtesy of Westinghouse Electric Corp.*)

beacon and radar patterns. The returns are corrected with signal processing so that they appear coincident on the display. Radiation patterns for the ARSR-3 beacon feed are shown in Figure 43-17.

Experimental beacon antennas using electronic scanning have been constructed. Cylindrical arrays are used to provide 360° azimuth scan. The version shown in Figure 43-18, built by Hazeltine Corporation's Wheeler Laboratory, mounts on the fixed portion of the tower for an S-band radar.¹⁴ The vertical pattern is a sector beam with 3 dB per degree horizon cutoff.

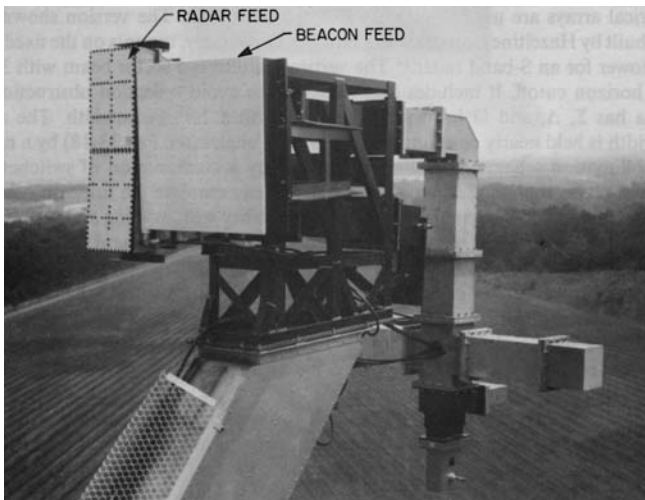


FIGURE 43-16 Integral beacon feed for the ARSR-3 (*Courtesy of Westinghouse Electric Corp.*)

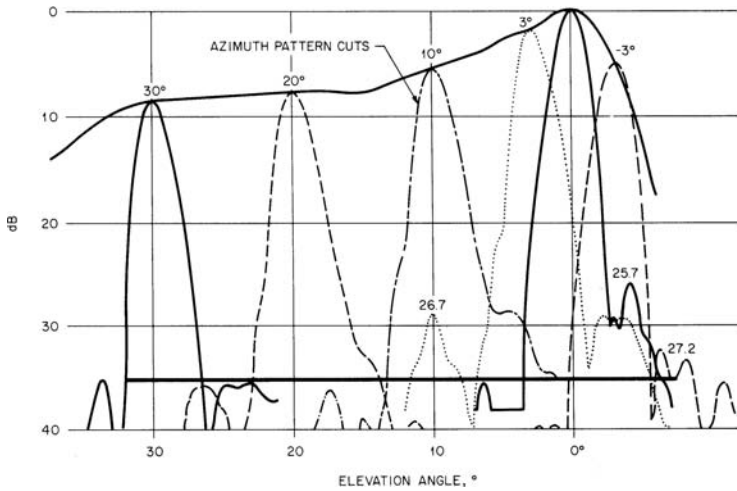


FIGURE 43-17 Typical radiation patterns for the ARSR-3 beacon antenna (*Courtesy of Westinghouse Electric Corp.*)

It includes beam hopover to avoid reflection obstructions. The antenna has Σ , Δ , and Ω horizontal patterns, with a 2.4° beamwidth. The azimuth beamwidth is held nearly constant for all elevation angles (see Figure 43-19) by a multiple-angle collimation scheme. Scanning, performed by a combination of switches (sector steps) and phase shifters (fine incremental steps), can emulate the uniform continuous motion of a mechanical-scan system (for compatibility with existing systems) or can be completely agile to support more advanced interrogation management schemes. To date, this antenna has not been widely deployed.

The array shown in Figure 43-20 has been developed for tactical applications.¹⁵ The 65° azimuth beam is scanned in increments of 5.6° using a modal Butler-matrix feed that

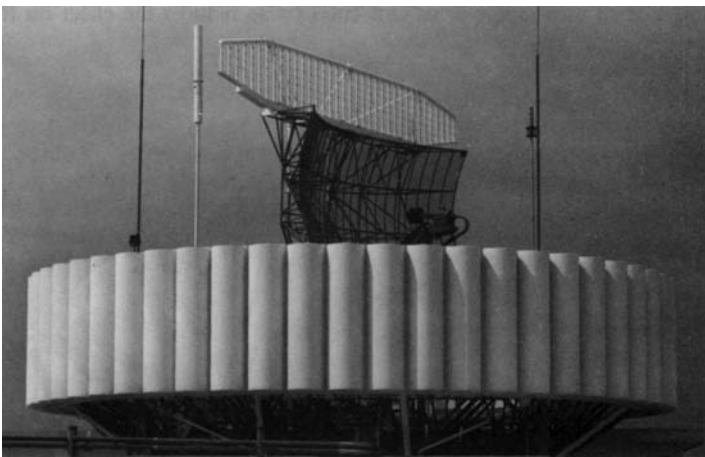


FIGURE 43-18 E-SCAN cylindrical array (*Courtesy of Hazeltine Corporation*)

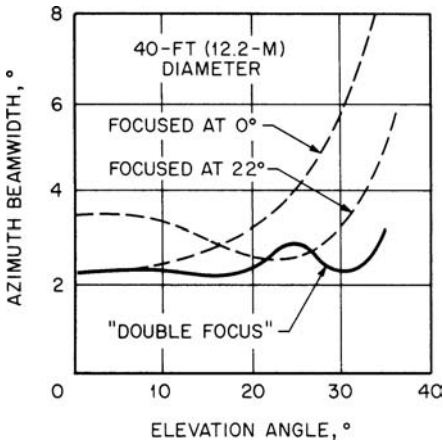


FIGURE 43-19 Beamwidth variation of E-SCAN array versus elevation angle (Courtesy of Hazeltine Corporation)

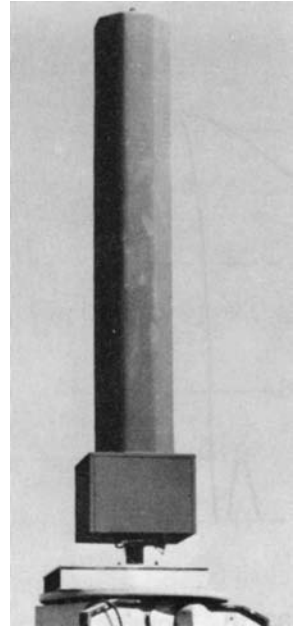


FIGURE 43-20 Tactical IFF-ATM cylindrical array (Courtesy of Bendix Communication Division)

excites all eight dipole columns. The SLS and monopulse functions share a difference-type pattern synthesized to provide backlobe coverage. All beams have the same elevation pattern with a high horizon cutoff rate.

The cylindrical array shown in Figure 43-21 has been developed for shipborne IFF use. The directional beam, complemented by a stationary omnidirectional SLS beam, is positioned by a Lockheed-proprietary trimode scanner system.¹⁶

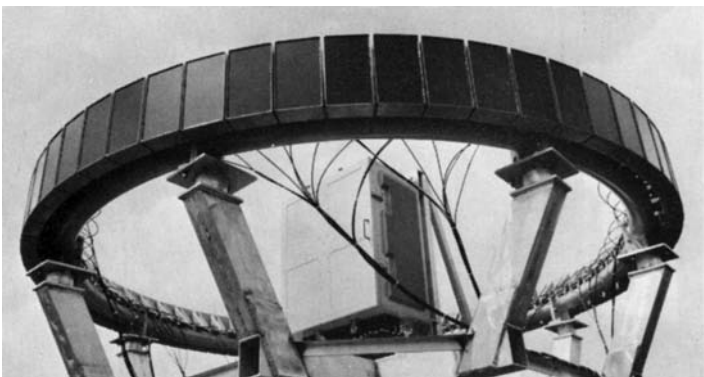


FIGURE 43-21 Shipborne ATC-IFF cylindrical array AN/OE 120/UPX (Courtesy of Lockheed Electronics Co.)

Airborne Systems

To date, most IFF interrogators on airborne platforms have been for military use. Airborne interrogators have not seen wide use in civilian aviation.

Airborne IFF interrogators almost always have limited availability of aperture size. They are typically used in conjunction with a radar at X- or Ku-band. Figure 43-22 shows a typical implementation in which L-band dipoles for IFF are mounted on the surface of a radar reflector. The dipoles use resonant techniques to reduce scattering at the radar frequency. Use of thin dipole arms and baluns also reduces the effect on the radar. Radar/beacon interaction is most easily controlled when the polarizations of the two systems are orthogonal. Because the beamwidth is always relatively large, beam sharpening using difference-pattern SLS is employed, with the power level for the difference pattern at nominally 9 dB above the sum pattern.

Other, similar radar antennas have used cavity-backed L-band slots instead of dipoles. The slot apertures are flush with the reflector surface, and the slots and cavities are carefully tuned to maximize reflectivity at the radar frequency. Folded $\lambda/2$ slots in which the slots are very thin (width $< 0.01\lambda$) have been found to be quite invisible to the radar.

IFF dipoles also have been added to “flat plate” radar antennas (planar arrays of waveguide slots). Thin, carefully tuned dipoles and orthogonal polarization are used to minimize interaction of the two systems.

The use of airborne interrogators on civilian aircraft has been investigated to provide traffic-alert and collision-avoidance information. An experimental antenna for this purpose is shown in Figure 43-23. This antenna can form either a directional beam or a notched beam in any one of eight azimuth directions. Using sidelobe suppression, the antenna is used to interrogate selected 45° sectors. The replies are received on multiple beams and are processed to provide 360° direction-finding capability with 8° rms accuracy. Top-loaded monopoles are used in this design to provide a low-drag antenna. Four monopoles spaced at approximately $\lambda/4$ are used. The dipoles are fed by a beam-forming network consisting of stripline couplers, diode switches, and phase shifters.

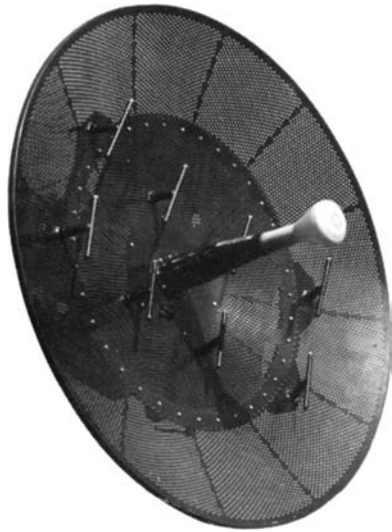


FIGURE 43-22 IFF array for airborne fire-control radar (Courtesy of Hazeltine Corporation)



FIGURE 43-23 Fuselage-mounted interrogator array for collision-avoidance system (Courtesy of Dalmo Victor Operations, Textron, Inc.)

REFERENCES

1. M. I. Skolnik, (ed.), "Beacons," *Radar Handbook*, Chap. 38 (New York: McGraw-Hill, 1970).
2. P. R. Drouilhet, "The Development of the ATC Radar Beacon System: Past, Present and Future," *IEEE Trans. Com.*, vol. Com-21, no. 5 (May 1973): 408-421.
3. D. Boyle, "DABS and ADSEL—The Next Generation of Secondary Radar," *Interavia* (March 1977): 221-222.
4. K. J. Keeping and J. -C. Sureau, "Scale Model Pattern Measurements of Aircraft L-Band Beacon Antennas," MIT Lincoln Lab. Proj. Rep. ATC-47 (FAA-RD-75-23), April 4, 1975.
5. G. J. Schliekert, "An Analysis of Aircraft L-Band Beacon Antenna Patterns," MIT Lincoln Lab. Proj. Rep. ATC-37 (FAA-RD-74-144), January 15, 1975.
6. J. E. Evans, "Synthesis of Equiripple Sector Antenna Patterns," *IEEE Trans. on Antennas and Propagat.*, vol. AP-24 (May 1976): 347-353.
7. A. R. Lopez, "Sharp Cut-Off Radiation Patterns," *IEEE Trans. on Antennas and Propagat.*, vol. AP-27 (November 1979): 820-824.
8. D. Karp and M. L. Wood, "DABS Monopulse Summary," MIT Lincoln Lab. Proj. Rep. ATC-72 (FAA-RD-76-219), February 4, 1977.
9. D. L. Sengupta and J. Zatkalik, "A Theoretical Study of the SLS and ISLS Mode Performance of Air Traffic Control Radar Beacon System," *Int. Radar Conf. Proc.* (April 1975): 132-137.
10. N. Marchand, "Evaluation of Lateral Displacement of SLS Antennas," *IEEE Trans. on Antennas and Propagat.*, vol. AP-22 (July 1974): 546-550.
11. M. C. Stevens, "Cossor Precision Secondary Radar," *Electron. Prog.*, vol. 14 (1972): 8-13.
12. P. Richardson, "Open Array Antenna for Air Traffic Control," *Texas Instruments Equip. Group Eng. J.*, (September-October 1979): 31-39.
13. C. A. Miller and W. G. Collins, "Operational Evaluation of an Air-Traffic Control Radar Beacon System Open Array Antenna," *Eascon Conf. Rec.*, Institute of Electrical and Electronics Engineers (1978): 176-185.
14. R. J. Giannini, J. Gutman, and P. Hanna, "A Cylindrical Phased Array Antenna for ATC Interrogation," *Microwave J.* (October 1973): 46-49.
15. J. H. Acoraci, "Small Lightweight Electronically Steerable Antenna Successfully Utilized in an Air-Traffic Management System," *IEEE Nat. Aerosp. Electron. Conf.*, vol. 1 (1979): 43-49.
16. D. S. Lerner, Power Distribution Network, April 17, 1973: U.S. Patent 3,728,648.

Chapter 44

Satellite Antennas

Robert B. Dybdal

The Aerospace Corporation

CONTENTS

44.1	INTRODUCTION	44-2
44.2	SPACE SEGMENT ANTENNAS	44-4
44.3	USER SEGMENT ANTENNAS	44-17
44.4	SATELLITE ANTENNA TESTING	44-21

44.1 INTRODUCTION

Antennas for satellite systems are required for both the space and user segments. Such antennas pose interesting design and test challenges. Space segment antenna designs are arguably the most diverse spacecraft technology varying with frequency, coverage requirements, and flexibility of operation. Likewise, user segment antenna designs vary drastically with frequency and data rate requirements, and can vary from very small designs as satellite services are extended to individual users to much larger ground terminals for high data rate services. The technology for both the space segment and the user segment is quite diverse and draws upon those technologies described in detail elsewhere.

Communication services are a major application for satellite systems and the diversity of antenna technology¹ that has been used for communication satellites has a significant range of performance capabilities. Both commercial and military services are commonly found and the differences in their objectives are apparent in the choice of technology. Commercial communication systems are strongly motivated to obtain as much communication capacity from their allocated bandwidth thereby increasing the amount of data that can be communicated. This motivation has resulted in the development of frequency and polarization reuse techniques. Frequency reuse uses the same bandwidth in portions of the satellite's field of view that are isolated by the satellite's antenna patterns. Polarization reuse techniques increase throughput by transmitting and receiving orthogonally polarized signals. Commercial systems are developed based on projected capacity demands in well defined coverage areas. Military systems, by contrast, require flexibility in their coverage characteristics in order to respond to changing geopolitical needs. Military services use technology to mitigate interference from undesired sources to protect authorized users. Other satellite applications provide navigation services and remote sensing data. The requirements for these applications provide differing challenges from communication applications. Finally, every satellite system requires tracking, telemetry, and control (TT&C) capabilities to determine the satellite's orbital location, provide health and status information on the satellite, and inject and validate commands to maintain satellite operation over its lifetime.

While specific applications have differing antenna technology requirements, these applications have common issues. The mantra for the space segment is size, weight, and power (SWaP) because of limitations imposed by the satellite launch capabilities. A significant development effort in composite materials and reduced weight fabrications has resulted. Space segment antennas must have the capability to withstand launch and on-orbit environments and consequently, the designs must be sufficiently rugged to survive launch as well as have adequate thermal protection for accommodating the on-orbit thermal extremes while having a low mass. Satellite antennas must also reliably operate unattended over the long orbital lifetimes. Reliability concerns result in rigorous prelaunch testing and the antenna designs are generally required to substitute redundant RF electronics to maintain operational capability over the required lifetime. In many cases, restrictions on the launch vehicle envelope result in the need for antenna deployment mechanisms to position antenna components to their operational configuration, so the reliability required for such mechanisms is understandably high. Still other requirements for electromagnetic interference/electromagnetic compatibility (EMI/EMC) are necessary to avoid mutual interference to/from satellite subsystems. The user segment also faces common requirements in their development/implementation. User antennas for communications have established requirements for antenna sidelobe control to avoid interference to/from adjacent satellites. The spectral crowding with other microwave services in adjacent frequency allocations, particularly at the lower microwave frequencies, requires increased design attention to protecting user systems from unintentional out-of-band interference from terrestrial users using other microwave services. As the trend to extend satellite services to individual users continues, design challenges include low fabrication cost to control the overall system acquisition

costs and developing reliable minimal testing techniques to assure design compliance on a production basis.

The satellite orbits provide other differences in space and user segment antenna designs. Predominately, satellites are in a geostationary earth orbit (GEO), which maintains these satellites in a fixed position above the earth as the earth rotates. This orbit locates satellites 20,000 nmi above the earth where the earth subtends about a 16.9° field of view. The combination of the long range separating the user segment and the satellite and the relatively narrow field of view dictates the use of high gain antennas on the space segment to permit link closure. Nonetheless, maintaining a fixed position over the earth has a significant advantage in providing continuous service to users. One disadvantage of this orbit is the limited coverage in polar regions. Applications, such as navigation, require complete global coverage including polar regions. These applications generally use inclined orbital constellations in a lower medium earth orbit (MEO) 12-hour orbit, with a 10,000 nmi altitude where the earth subtends about 28.7° . Other applications, particularly for remote sensing, use still lower low earth orbit (LEO) inclined 450 nmi orbits where the earth subtends 83° . This is used both to meet resolution requirements in their determination of feature contrasts and to achieve global coverage. As the satellite altitude decreases, the earth subtends a wider field of view from these lower altitude orbits, while a broader antenna coverage is required for the space segment. Satellite orbits between MEO and LEO altitude extremes are less attractive because the ionizing radiation from the Van Allen belts imposes significant protection requirements for electronic components. These orbital alternatives also impact the user segment. For example, geostationary satellites that maintain their position have very minor user antenna tracking requirements, especially in comparison to low altitude satellites that require user terminals to follow the satellite's trajectory when the satellite is in view. The reduced range to LEO satellites can also reduce user performance requirements. However, since an individual satellite within the constellation has a limited time in view of the user, switching to another satellite within the constellation is necessary to maintain service availability.

Clearly, antennas for satellite systems have a wealth of challenges for design, development, and operation. The subject is truly multidiscipline, requiring not only an RF understanding but also mechanical and thermal awareness, interface issues with RF and control electronics, weight, volume and power, implementation costs, and extensive testing requirements following both established practice and test issues peculiar to space systems. The diverse technologies used in satellite systems provide opportunities for creative designs, and, as in the past, significant development challenges exist for future satellite system antennas.

At a system level, two parameters specify antenna system level performance for satellite communications:

1. The G/T (antenna gain, G , divided by system noise temperature, T) for receiving antennas
2. Effective radiated power (ERP), equal to antenna gain multiplied by transmitted power, for transmitting antennas

The required values of these system parameters are derived from link analyses for communications. Generally, E_b/N_o , the received signal energy per bit divided by the noise spectral density, is determined from

$$E_b/N_o = \text{ERP } G/T [L/(kR_b(4\pi R/\lambda)^2)]$$

where L is the system loss that includes propagation loss and receiver implementation loss, k is Boltzmann's constant, R_b is the bit rate of the data, R is the separation between transmitting and receiving antennas, and λ is the wavelength corresponding to the data rate. The value of E_b/N_o is determined from the data fidelity requirements in terms of bit error

rate (BER), the statistical average of the number of bits received in error out of a specified total number. The BER values are a function of E_b/N_o , which is commonly referred to as a “waterfall” curve that is dependent on the modulation format used and the receiver’s implementation loss. Both ERP and G/T directly depend on the antenna gain value.

The space segment is characterized by the uplink G/T for the receive antenna and the downlink ERP for the transmit antenna. The system parameters for the space segment are generally specified as minimum values over their required coverage area and also include the effects of antenna pointing errors resulting from the satellite’s attitude uncertainties. The user segment is characterized by the uplink ERP transmit capability and the downlink G/T receive capability and are burdened by user antenna pointing errors.

The system noise temperature T has two components, the antenna temperature and the receiver noise temperature of the electronics. Both the antenna gain and system noise temperature must be specified at a common reference plane (for example, the input terminals of a low noise amplifier) to quantify the G/T capability of the design. For the space segment, the specification of earth-directed antennas typically references an ambient 290 K noise level for the antenna temperature. In practice, antenna temperature values² are somewhat lower, resulting from the emission characteristics of the particular portion of the earth the antenna is servicing. The antenna temperature is typically higher than existing low noise receivers and it limits the uplink system sensitivity. An exception is crosslink antennas that provide connectivity between satellites; these antennas view a cold 3 K background noise temperature. The user antenna temperature, as will be discussed in Section 44.3, depends on atmospheric path loss that is a function of the elevation angle to the satellite and ohmic loss.

44.2 SPACE SEGMENT ANTENNAS

One approach in describing space segment antennas separates the designs by their coverage requirements. This separation results in earth coverage designs that span the earth’s field of view, spot coverage designs that serve a limited portion of the earth’s field of view, multiple beam designs that serve the entire or a limited portion of the earth’s field of view, and hemispheric coverage antennas needed for TT&C applications. When connectivity between satellites is required, point-to-point antennas are used in crosslink subsystems. Such crosslink subsystems allow global connectivity without the need for a relayed link with a ground terminal. Additionally, antennas for remote sensing applications require narrow beamwidths to obtain the required resolution for extracting the desired parameters and these narrow beamwidth antennas must be scanned to obtain their coverage requirements. This separation by coverage will be used in this discussion.

Earth Coverage Designs

The simplest satellite antenna design provides coverage over the complete earth disk visible from the satellite. Earth coverage antennas were the only feasible option for early satellites because of weight constraints and satellite attitude stability. Today, such designs are used in a limited number of applications, such as user access and routing, gateway (a link to a ground terminal that processes the uplink information into the downlink), services to avoid steering a more directive antenna beam, and so on. The technology for geostationary satellites generally uses simple horn antennas at microwave frequencies; corrugated or dual mode technologies are commonly used to obtain efficient rotationally symmetric patterns with good axial ratio performance. Such antennas are both small and lightweight. When both receive and transmit earth coverage services are required, and

a tradeoff exists between using a single horn for both receive and transmit or using individual horns for each function. When the individual horns are used, physical separation provides isolation to reduce the filtering requirements for diplexing and, with reduced filtering requirements, lower loss can be achieved.

When earth coverage antennas are used in orbits lower than geostationary, the earth subtends a greater angle and the range differences between the subsatellite point and the earth's limb also increase. In this case, antenna designs with a shaped pattern can be used to equalize performance for users within the visible portion of the earth. Arguably, the best known example of such designs is the antenna array³ used in navigational satellites. This array design produces higher gain towards the earth limb than at the subsatellite point, in order to offset the range differences between the subsatellite point and the limb of the earth.

Spot Coverage Designs

Spot coverage antennas serve a limited portion of the available field of view. A simple example is a parabola antenna mechanically positioned to serve a specified region of the earth. If communication needs change, the antenna can be repositioned to provide coverage in that area. Since a limited portion of the field of view is covered with a narrow antenna beamwidth, the antenna coverage area must also encompass the uncertainty resulting from the attitude stability of the spacecraft.

In some applications, coverage over the specified area and reduced coverage in other portions of the field of view is required. Such requirements typically arise to satisfy regulatory issues that limit coverage in other geographic areas or when the same satellite reuses the same frequency subband in different portions of the earth field of view and subsequent isolation between the two coverage areas is required to avoid cochannel interference. Often, the single beam parabola can satisfy the coverage requirements, but the sidelobes from this design may not adequately suppress the radiation beyond the specified coverage area. What is needed in such cases is an antenna design which has a sufficiently large aperture and a corresponding sidelobe response with the angular resolution to control its coverage beyond the specified area.

One means for achieving both the coverage and reduced sidelobe levels beyond the coverage area uses a parabola with a cluster of feeds in the focal region, producing a corresponding cluster of beams. The individual beams in this cluster are combined to produce the specified coverage area and the antenna pattern rolloff beyond the coverage area is dictated by the sidelobe response of the large aperture. The combining can be passive and implemented with fixed power dividers or active where each element has an active RF device and the collection is combined to form the coverage area. Active combining imposes amplitude and phase tracking requirements on the electronic devices, in order to maintain coverage characteristics. In addition, the coefficients that are used in the active combining can also be varied to change the antenna pattern coverage during the satellite's lifetime. This operational flexibility is obtained at the expense of increased weight and power consumption. The variation in the active combining is commanded through the satellite's TT&C subsystem. In some cases, the specified coverage area has an irregular shape and, by combining a large number of individual beams, the overall coverage can closely conform to such shapes. Typical examples of such requirements are coverage for time zones or coverage restricted, for regulatory reasons, to particular political boundaries. As the number of narrow beams increases to cover the specified area, the antenna aperture size has a corresponding increase to produce the narrow beamwidths. A better fit for the desired coverage irregularities can be achieved, but the spacecraft must accommodate this larger aperture size and more complex feed network.

Another approach to providing irregular coverage areas shapes the reflector surface.⁴⁻⁶ This approach still requires a larger aperture than would be used by a single spot beam; the required aperture size is comparable to multiple feed cluster designs so that the angular resolution and sidelobe rolloff requirements can be met. The design does simplify the feed requirements and typically a single feed horn is used. Synthesis approaches have been developed for this shaping technique. The pattern of this antenna design is fixed by the reflector surface shaping and unlike active feed combining cannot be changed by command.

Commonly in commercial system designs, the communication capacity becomes limited by the available bandwidth in the frequency allocation. The communication capacity to spot coverage areas can be increased by a factor of two when independent data signals are communicated by using orthogonal polarizations (a technique referred to as polarization reuse). Independent data streams are communicated on each polarization component; a technique used by direct broadcast services. In operation design, attention must be paid to maintaining isolation between the orthogonally polarized channels to avoid cochannel interference, as the design of both the space and user segment antennas requires high purity in the orthogonal polarizations. A typical requirement for space segment antennas is a -27 dB cross polarization level, and similar polarization purity requirements are levied on user antennas. These polarization purity requirements are specific for particular system applications. The isolation provided by orthogonal polarization depends on the polarization purity of both the incident field and the receiving antenna. Often, polarization is expressed in terms of axial ratio, defined as the ratio of the major and minor axes of the polarization ellipse. The isolation depends on the axial ratios of the incident field and receiving antenna, as well as the relative orientation of their polarization ellipses. In many cases, the axial ratio is known but the relative orientations of the polarization ellipses are generally unknown.

The statistics of the isolation⁷ have been derived assuming the axial ratio values of the incident field and the receiving antenna are known, but their polarization ellipse orientations are unknown. The unknown orientations of the polarization relative to one another can have values between 0° and 180° . The statistics are derived by assuming their values are equally likely and uniformly distributed over that range of relative orientations. The minimum, maximum, mean, and rms spread of the isolation may be evaluated from the axial ratio values of the incident field and receiving antenna. The statistics are derived from the polarization efficiency. When the incident field and receiving antenna have opposite polarization senses, the polarization efficiency is the isolation; when the incident field and the receiving antenna have the same polarization sense, the polarization efficiency is the polarization mismatch loss. The polarization efficiency is given by

$$\eta_p = (1/2) + A + B \cos \Delta$$

where

$$A = \frac{2r_w r_r}{(1 + r_w^2)(1 + r_r^2)}$$

$$B = \frac{(1 - r_w^2)(1 - r_r^2)}{2(1 + r_w^2)(1 + r_r^2)}$$

and r_w and r_r are the axial ratios of the incident wave and the receiving antenna, respectively. When the incident field and the receiving handedness is the same, A is positive, and when their handedness is opposite, A is negative. The angle, Δ , is the phase difference between right- and left-handed components of the polarization ellipses and equals twice the difference in the tilt angle alignments of the ellipses. The commonly used bounds on polarization

efficiency are $(1/2) + A \pm B$. Assuming the tilt angle differences are equally likely and uniformly distributed from 0 to 180°, the first order (mean) statistics become

$$E_p = (1/2\pi) \int_0^{2\pi} \eta_p d\Delta \\ = (1/2) + A$$

When the incident field and receiving antenna have the same sense, the mean efficiency is $\geq 1/2$, and when the handedness is opposite, the efficiency is $\leq 1/2$ yielding the polarization isolation. The second order (variance) statistics become

$$V_p = (1/2\pi) \int_0^{2\pi} (\eta_p - E_p)^2 d\Delta \\ = B^2/2$$

so that the standard deviation equals $B/\sqrt{2}$. Since these statistics have a non-zero mean value, the second order statistics are generally expressed as $\pm 1\sigma$ about the mean value, also known as the rms spread about the mean. Notice that the polarization efficiency statistics are not Gaussian. The peak to peak bounds are $\pm B$ for these statistics while the rms spread is $\pm 0.707 B$. Thus, the bounds equal 1.41 times the standard deviation.

Example statistics for isolation and polarization mismatch loss in Figure 44-1 illustrate the variations, assuming an incident field with a 1 dB axial ratio and receiving antennas with axial ratio values as great as 4 dB. The isolation has a finite value even with an ideally polarized receiving antenna resulting from the cross polarization of the incident field and also has a possibility to achieve ideal cancellation when received by an orthogonal polarization, 1 dB axial ratio receiving antenna. The polarization mismatch has minimal values for the range of receiving axial ratio values and again, zero loss is probable with an ideally matched 1 dB axial ratio receiving antenna. It is also apparent these statistics do not have a Gaussian distribution. This example illustrates that achieving high values of polarization isolation require significant design attention to the polarization purity of the link antennas and minimizing depolarizing effects in the link.

Military communication systems generally have spot coverage requirements for serving military theaters and this coverage is mechanically repositioned to respond to changing geopolitical requirements. Interference threats and issues of privacy are also concerns to military systems. These concerns are addressed through spread spectrum modulation formats, error correction techniques, data interleaving, and encryption, as well as the protection that can be achieved through sidelobe control and adaptive antenna processing⁸ to negate

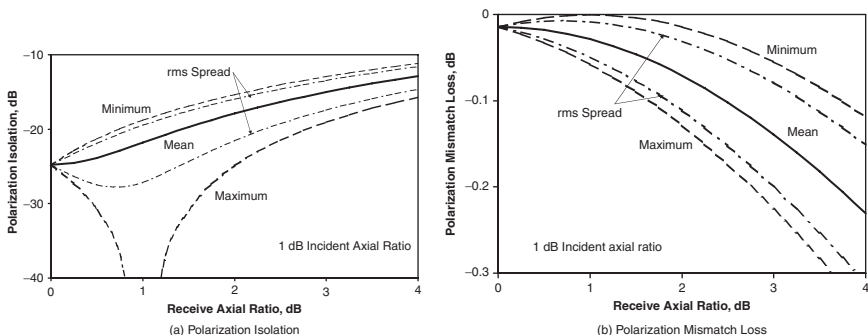


FIGURE 44-1 Polarization statistics (after R. B. Dybdal⁷ © IEEE 1999)

uplink interference. Sidelobe protection reduces interference power beyond the design coverage area. Adaptive processing using a cluster of uplink beams with a reflector design⁹ provides the ability to negate interference within the coverage area. Reflector antenna technology,¹⁰ for adaptive uplink antennas, is generally preferred to array technology both for performance reasons and SWaP. Rather than using a fixed combination of beams to conform to a specified coverage shape, the adaptive design provides amplitude and phase control to dynamically produce antenna pattern nulls in the direction of interference sources. Because these systems generally use spread spectrum modulation as well, for interference protection, the interference power must significantly exceed the user power if the interference is to be effective.

Adaptive antennas must:

1. Distinguish desired signals from interference
2. Respond to interference that can degrade system performance
3. Combine antenna elements with adaptive weighting circuitry to reduce interference
4. Provide a control processor to derive the adaptive weight values to satisfy an optimization criteria
5. Monitor and respond to changes in the interference.

Since spread spectrum techniques are generally used, interference is distinguished from desired signals on the basis of the spreading technique. The adaptive design is generally activated when a threshold power level is exceeded, indicating interference signals that are sufficiently strong to exceed the capabilities of other interference protection techniques. Correlation techniques are used to determine the amplitude and phase weighting of the individual beams. Iterative correlation measurements are typically processed by an algorithm based on least mean square measures. This algorithm is derived to satisfy an optimization criterion such as maximum $S/(N + I)$ or signal to noise plus interference ratio.

An interference scenario is specified to guide the adaptive antenna development and defines the number, geographic distribution, and RF performance of the interference sources. Adaptive system design¹¹ proceeds by developing a simulation program based on a candidate RF design, the RF circuitry, and the correlation and control electronics associated with the design. Two types of performance measures are required, one specifying the transient time in responding to interference initiation and the second measuring steady state performance. When nulls are produced within the coverage area, portions of the coverage area have insufficient performance to permit link closure for the users. The percent coverage area is a quantitative measure of performance and is defined as the fraction of the original design coverage area remaining after interference cancellation that permits user operation. A threshold $S/(N + I)$ is determined where acceptable communications as measured by a BER value can occur. This threshold $S/(N + I)$ level provides a means to define a coverage boundary and hence determine the percent coverage area. These measures depend on specific interference source locations. The simulation program varies the interference locations, power levels, spectral contents, and so on, as defined by the scenario in a Monte Carlo manner, in order to obtain statistical answers to the transient time and percent coverage characteristics as needed to evaluate adaptive system performance. Selected cases that are representative of the scenario are used to develop a test program, since it is impractical to test on a Monte Carlo basis. These test cases are then used to validate the simulation and the statistical measures of performance are obtained from the validated simulation.

The minimum angular separation between users and interference sources is an important parameter for adaptive uplink antenna design. Statistical answers for this parameter

can be obtained from the adaptive system simulation. The separation⁹ depends on the minimum beamwidth that can be formed by the antenna, the performance above a threshold level for a user, and the incidence angle on the earth's surface. An estimated separation distance in Figure 44-2 illustrates these factors as a function of the minimum beamwidth of the aperture and is based on aligning the antenna pattern null between the main beam and first sidelobe with the interference source. These systems, particularly at the higher EHF frequencies, are margined to maintain link availability during rain, and hence, under clear conditions, the margin can be reduced in the interest of interference negation and still maintain operation at the threshold $S/(N + I)$. Two values of antenna gain loss, 3 and 10 dB, are illustrated in Figure 44-2; if 10 dB of margin loss can be tolerated, the system can maintain threshold performance closer to the interference than if only 3 dB of margin loss can be accommodated, for example. Finally, the minimum separation differs between the subsatellite point and a lower 20° elevation angle because the beams spread further over the earth's surface at the elevation angles. These values provide a first order estimate of the required aperture size for meeting the desired angular resolution between users and interference, and are supported by more detailed simulation results. A further parameter of interest is the achievable amount of interference suppression. Interference cancellation is inherently a subtraction process, and the cancellation depth depends on the ability to match the amplitude and phase responses in the adaptive channels over the required bandwidth. The tolerance requirements in Figure 44-3 become increasingly stringent as the required interference suppression increases.

The protection of high data rate services¹² in a military area benefits from using a high gain space segment antenna and from the operation of EHF frequencies with wide bandwidth frequency allocations. Antenna pointing must be maintained and the attitude stability of the satellite becomes a significant contributor to the antenna pointing loss. One solution¹³ lies in actively tracking a ground beacon to maintain pointing; for this a coded beacon signal and correlation processing on the satellite can be used. This technique can also be used to reduce interference. The narrow beamwidth limits interference from ground-based sources removed from the ground terminal location. Because system transfers high data rate signals, only limited interference protection is afforded by spread spectrum modulation. Protection from interference located close to the ground terminal is provided by the narrow beamwidth.

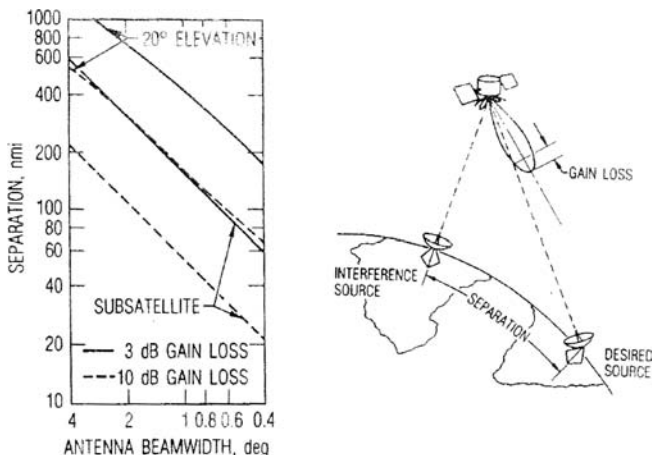


FIGURE 44-2 Interference/User angular separation (after K. M. Soo Hoo and R. B. Dybdal⁹ © IEEE 1989)

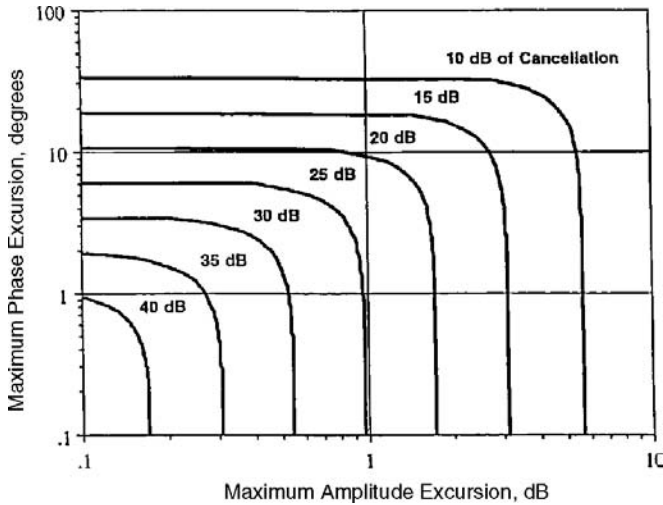


FIGURE 44-3 Amplitude and phase tolerance for adaptive antennas

Additional protection can be obtained by offsetting the satellite antenna from the ground terminal in a direction away from the interference source, as illustrated in Figure 44-4. The ratio of the signal to interference power increases from R to R' as indicated as the beam is repositioned, with some loss in the received S/N . Because a significant rain margin is typically required for link availability, some loss in peak gain can generally be sacrificed in order to gain protection from interference close to the ground terminal. As the beam

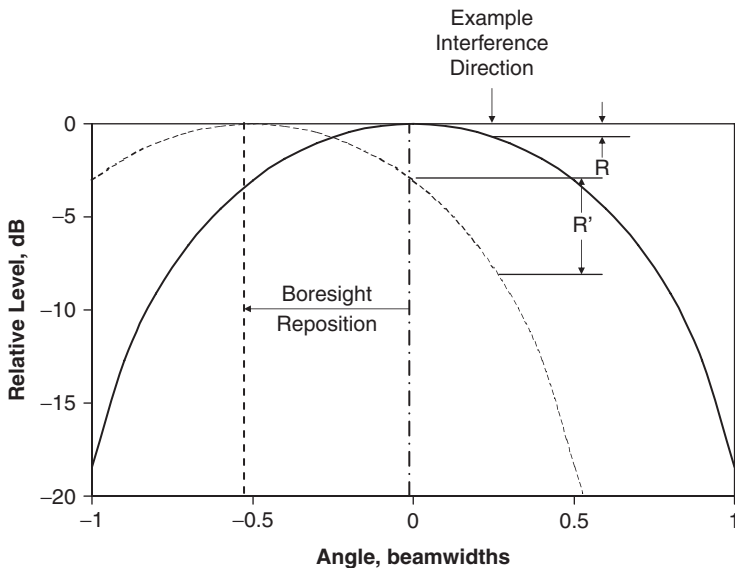


FIGURE 44-4 Adaptive beam repositioning illustration (after R. B. Dybdal and S. J. Curry¹³
© IEEE 1996)

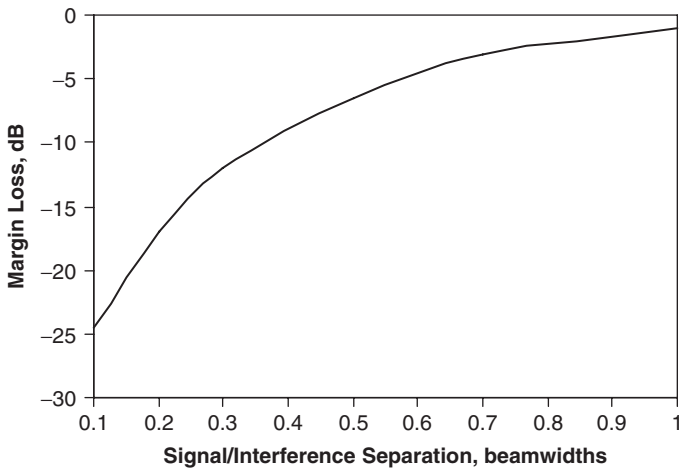


FIGURE 44-5 Margin loss with beam repositioning (after R. B. Dybdal and S. J. Curry¹³ © IEEE 1996)

is offset from the ground terminal away from the interference source, the interference is reduced more than the desired signal because of the increased slope of the antenna pattern. As the signal margin permits, it is desirable to offset the narrow beam in order to align the null between the main beam and the first sidelobe with the interference source to improve the $S/(N + I)$. The required margin is shown in Figure 44-5 as a function of the angular separation between the user and interference source.

The implementation of adaptive beam repositioning follows the functional diagram in Figure 44-6. The desired signal is accompanied by a pseudo-random code for tracking purposes. The correlation of the data (Σ channel) and the tracking (Δ channel) signals produces outputs for tracking that are not affected by interference, because the interference signal is not correlated with the coded desired signal. Thus, independent of fluctuations in the desired received signal level that may result from weather events, the desired signal location is always determined. When interference is present, the cross correlation of the data

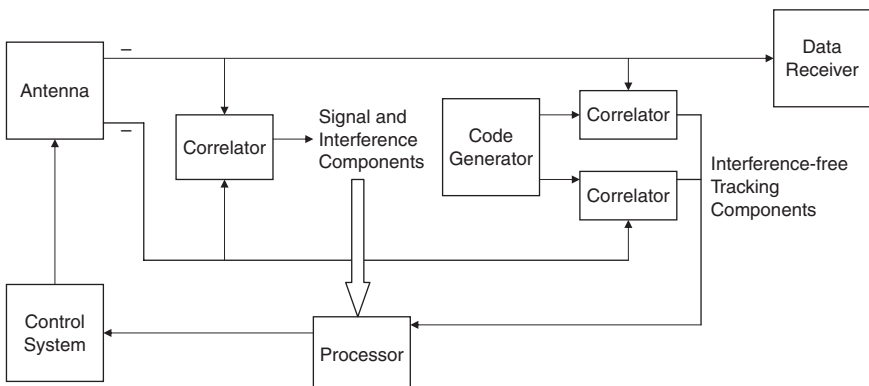


FIGURE 44-6 Functional diagram of adaptive beam repositioning (after R. B. Dybdal and S. J. Curry¹³ © IEEE 1996)

and tracking channels contains components that include both the desired signal and interference; additionally, power meters on the data and tracking channels can also indicate interference initiation. The cross correlation output level indicates the interference strength and open loop commands can be initiated to reduce the interference as indicated in the cross correlation output. During such commanding, the desired signal level in both the data and tracking channels can be monitored independent of the interference levels, and the signal reduction relative to the margin and the signal growth in the tracking channel can be observed.

Antennas for spot beam applications are typically reflector technology, since a limited field of view is required and reflector antennas have attractive SWaP. In many cases, the number of antennas that can be used is limited by the available space on the satellite. One approach¹⁴ for systems using multiple frequency bands integrated several such bands into the same aperture to provide more compact designs. This design example provides both shaped and multiple beam capabilities. In many cases, frequency selective surface technology can be used to isolate frequency bands in multiple frequency designs. More commonly, the coverage requirements must be maintained for both the uplink and downlink services that use different frequency bands, the required aperture sizes for the uplink and downlink antennas differ, and separate uplink and downlink antennas are often used to maintain the same coverage characteristics at both frequencies to reduce design complexity.

Multiple Beam Antennas

In many applications, the system design requires covering large portions of the field of view with high G/T and ERP beams for link closure reasons and/or achieving greater data capacity than the frequency allocation can support, if a single coverage beam is used. Multiple beam antenna technology provides the means to meet these objectives for such applications. The field of view is served by a collection of beams where the narrow antenna beamwidth is used to meet G/T and ERP requirements. Isolation between individual beams permits frequency subbands to be reused in separated coverage areas and this frequency reuse increases the system data capacity.

Multiple beam antennas typically use reflector antennas with a cluster of feed elements in the focal region to generated independent beams in different directions. The beams are arranged on an equilateral triangular grid. Commonly, a central beam is surrounded by a ring of six feeds surrounded again by a ring of 12 feeds, then by a ring of 18 feeds, and so on. This results in the sequence, 7, 19, 37, 61, 91, ... and so on for the number of beams to service a coverage area. Figure 44-7 illustrates an example beam pattern for 91 beams. The required number of beams greatly increases as the beamwidth of the individual beam decreases, and the peak gain for each beam also increases. Figure 44-8 illustrates a typical variation of these parameters for the complete field of view coverage of geosynchronous satellites, along with the required aperture size at selected frequencies. The required aperture size variation is also indicated in this figure for selected frequencies. As the beamwidth of the individual beams becomes narrower, not only is the system complexity increased by the number of beams, at lower frequencies, deployment of the reflector surface is also required at lower operating frequencies.

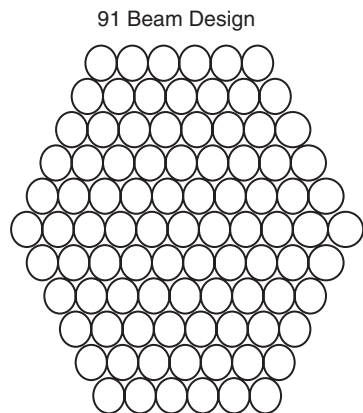


FIGURE 44-7 Multiple beam arrangement for 91 beams

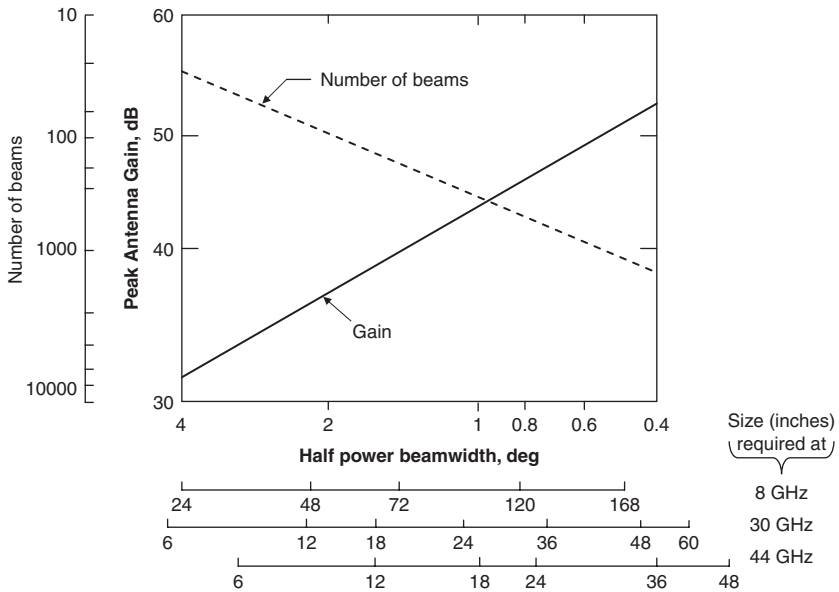


FIGURE 44-8 Multiple beam parameters

A design conflict¹⁵ results between the desire for beams with low sidelobes to isolate beam positions and the desire for the beams to crossover at a high pattern level to maximize the minimum gain within the coverage area. Reduced sidelobe levels require an aperture distribution with amplitude tapering that in turn requires directive antenna feeds. However, the antenna feed size is limited by the inability for the antenna feeds to physically overlap. Two different approaches have been investigated for addressing this conflict between beam overlap and sidelobe levels. One approach¹⁶ defocuses the feed cluster and under-illuminates the reflector. A second approach¹⁷ develops low sidelobes by combining a feed with a weighted sum of the six adjacent feeds, which results in a complex beamformer. The scan performance of reflector antennas¹⁸ is limited by phase aberrations for feed positions removed from the reflector's focal point. Reflectors with a long f/D ratio and dual reflector designs benefiting from the magnification ratio are means for reducing the phase aberrations and producing high quality beams over a field of view.

Multiple beams commonly use frequency reuse techniques to extend system capacity. Frequency reuse imposes isolation requirements between beam positions that share a common frequency subband. The isolation between beam positions depends on the sidelobe response of the closest beams sharing the same subband. The required isolation depends on the susceptibility of the modulation to cochannel interference and the dynamic range of the users. In practice, these two factors are used to devise a frequency reuse plan defining the subband assignments to individual beams within the multiple beam collection. Since the beams are typically arranged on an equilateral triangular grid, the minimum number of subbands to avoid the same subband in adjacent beam is three. More commonly, a frequency reuse plan based on seven frequency subbands is selected to provide adequate isolation.

Multiple beam antennas can also be used in a limited portion of the earth's field of view¹⁹ to provide high data rate, high capability communications to a limited geographical area. As the individual beamwidths of the design become narrower, their width can become

comparable to the attitude stability of the spacecraft. In such cases, actively tracking ground beacons²⁰ can be used to maintain beam pointing. Such designs are attractive for military communications to theater operations. Not only are the high data capacity advantages provided by such designs, protection from interference is afforded by the narrow beamwidths and the effectiveness such interference is limited to a few beam positions.

In multiple beam operation, capacity demands are often not uniformly distributed over the collection of beams and can dynamically vary. One approach²¹ to provide high capacity in a limited number of beam positions would use a frequency reuse plan providing a basic capacity to the individual beams within the collection along with separate frequency subbands that can be switched to beam positions with higher capacity demands as required. Because these frequency subbands do not overlap the subbands in the basic service, the necessary isolation is provided. As memory technology expands and becomes space qualified, storing non-time critical data for transmission at a later time when capacity is available may also provide a means to lower the peak capacity requirements. While such approaches add flexibility and capacity to the system design, they also add complexity and the need for network control.

The preponderance of multiple antenna technology for GEO satellites is reflector designs because the earth subtends a relatively small field of view. When multiple beam designs are used at lower microwave frequencies, the required aperture size requires the use of deployable reflector technology. When multiple beam designs are used for satellites in lower LEO orbits, beam generation over a wider field of view is required. For such applications, array technology is used to produce relatively broad beams over the larger field of view. Such array designs²²⁻²³ have a manageable number of elements, whereas applying array technology at the higher GEO altitudes results in an excessive number of elements. Array designs require active receive and transmit elements to offset losses in their corporate feed networks resulting in a prime power penalty. Transmit arrays require attention to the transmitter operating points because intermodulation of carrier components when multiple beams are routed through the transmit modules degrade performance. Failure of active elements has a minor effect on the peak gain levels of array antennas, but the sidelobe response²⁴ does not have the same graceful degradation. If a small number of elements fail, rephasing can maintain sidelobe performance, but as the number of failed elements increase, rephasing cannot maintain sidelobe performance. Further attention is also required to devise on-orbit calibration techniques to identify and verify array element excitation errors. Design attention is required to establish performance projected to the end-of-life levels accounting for the anticipated number of element failures. Further, the array elements must be commanded to produce the required beam patterns and beam steering, and such control also incurs additional weight and power impacts. As digital technology matures, digital beamforming techniques will evolve. Development experience for this technology²⁵ is described for a multiple beam design, albeit for a mobile user terminal.

Crosslink Antennas

Communication links between satellites provide a means to relay information between system users that are not in the field of view of a single satellite, and are referred to as *crosslinks*. Crosslinks permit global linkage of users without using a relay through ground stations. The antennas used in crosslinks are point-to-point configurations in contrast to other spacecraft antennas that serve a coverage area on the earth's surface.

Antennas for crosslink applications are typically reflector designs mechanically pointed at the corresponding antennas on other satellites within the constellation. Such designs use antenna tracking²⁶ to maintain beam alignment with their destination satellite. For GEO satellites, the antenna must have the capability to track other satellites within the equatorial plane and accommodate the elevation variations that depend on how well the satellite is stationkept; at GEO, orbital perturbations result in satellites following a figure eight motion

about the equatorial plane over a 24-hour period. Spacecraft thrusters control the size of this figure eight maintaining the satellite's position on the equatorial plane. This adjustment of the satellite's orbit is referred to as stationkeeping. Crosslink antennas, therefore, are required to provide principally azimuth positioning to link with different satellites within the constellation and a smaller elevation positioning requirement dependent on how well the satellites maintain stationkeeping.

Crosslink systems often operate at 60 GHz where wide bandwidth allocations exist to transfer high data rates. The 60 GHz frequency range corresponds to the resonances of oxygen absorption spectra^{27,28} and provides broad bandwidth, high level attenuation of terrestrial signals. Thus, operation at 60 GHz also isolates the crosslinks from ground based interference.²⁹ Satellite crosslinks often transfer high data rate data between satellites. The EHF operation also results in antennas with a compact physical size—an important factor in satellite systems. These antennas achieve high gain performance and their corresponding narrow beamwidths require antenna tracking capabilities to maintain beam alignment with adjacent satellites.

TT&C Antennas

All satellites require a TT&C subsystem to provide information on the satellite's orbital position, monitor the health and status of the spacecraft, and to inject and validate commands for the satellite's operation. The spacecraft's TT&C subsystem communicates to a dedicated mission control station that maintains the operation of the satellite. Antenna systems for TT&C applications require two types of antenna coverage. During launch operations, complete spherical coverage is required so if the spacecraft begins to tumble, a command can be injected to attempt to control the spacecraft independent of the satellite's orientation. Once on orbit, an earth coverage pattern is required so that alternative mission control stations can perform the TT&C operations. However, because the data rate requirements for TT&C applications are modest and ground terminal designs with ample performance can be configured, many satellite programs simply use the earth facing hemispheric coverage antenna to reduce design complexity.

Two antennas that provide hemispheric coverage are typically used to satisfy the spherical coverage requirement. One antenna covers the forward hemisphere and is also used in the on orbit position, and the second covers the rear hemisphere. During launch operations, the antenna with the higher signal power is used. Attempting to coherently combine the two antennas results in significant pattern ripple where their patterns overlap because of the physical separation between the two antennas. These broad hemispheric coverage requirements have the generic problem of isolating the TT&C antenna from the satellite blockage and scattering. Such systems typically operate at lower microwave frequencies with a relatively long wavelength. The long wavelength further limits the ability to control antenna backlobes that interact with the satellite's structure with a compact design. The receive and transmit frequencies can also be a relatively large percentage bandwidth. To satisfy these diverse requirements, generally a frequency independent antenna is used and the antennas are mounted on booms to provide a clear field of view and to minimize the interactions with the spacecraft structure. Design attention to reducing the backlobes of TT&C antennas³⁰ and the development of analysis and measurement techniques for projecting and verifying the performance of TT&C antennas in the presence of the spacecraft structure would be useful capabilities.

Remote Sensing Antennas

Remote sensing satellites can use RF sensors³¹ to collect data. Two types of sensors have been developed: active or radar sensors and passive or radiometric sensors. Such sensors

are used in low orbiting satellites for global coverage characteristics and high resolution of surface features.

Radar sensors³² for remote sensing applications use array antennas operated in a synthetic aperture mode. An example design is the SIR-C radar flown on the space shuttle. This design uses two active array antennas, one operating a L-band and the other operating at C-band. Additionally, the program also uses an X-band slotted waveguide array. Such a design approach allows the evaluation of image quality at different frequencies. The arrays are also capable of operation using orthogonally polarized beams so that the polarization contrast can be judged. Like other spaceborne antennas, SWaP has a significant impact on the design. Further development of this technology can be anticipated in the future; as the remote sensing data gathered by this effort provides a proof of concept and an example data base with which to refine estimates of remote sensing extraction algorithms.

Two categories of radiometric designs are used in passive RF remote sensing applications.³³ One category uses narrow antenna beamwidths that are mechanically scanned to obtain data across a terrain swath beneath the satellite. The beam is scanned across the swath and the spacecraft orbital trajectory fills in the swath in subsequent scans. A portion of this scan looks at reference temperatures to calibrate the radiometer, a “hot” reference is provided by an ambient temperature and a “cold” reference is provided by the deep space background with a 3 K temperature.

Radiometric sensors receive data from the emission of the surfaces and contrasts of the surface emission properties allow the generation of an image. Such sensors normally look obliquely at the earth’s surface and the combination of the polarization dependence of the surface emissivity and multiple frequency data allows estimates of remote sensing parameters. The resolution of the emission temperatures depends on the ΔT of the radiometer that is defined by

$$\Delta T = T_s / (B\tau)^{1/2}$$

where T_s is the total system temperature, B is the IF bandwidth, and τ is the integration time. Physically ΔT is the standard deviation of the output temperature. For specification purposes of the radiometric sensitivity, a 290 K antenna temperature is used as a reference value and the integration time is 1 second. In practice, the resolution of the emission temperature depends on the integration time at each pixel (picture element). Design attention to minimizing RF loss and to utilizing low noise amplifiers is necessary to achieve good resolution in emission temperature as measured by a low ΔT .

The second performance parameter for radiometric system concerns the spatial resolution of the design. The spatial resolution is measured by the antenna’s solid beam efficiency. The solid beam efficiency is the ratio of the integrated antenna pattern over a specified solid angle for a particular polarization to the total integrated pattern for all polarizations. The sense of this definition is to examine the collected emission over a specified solid angle in comparison to the emission collected over the entire pattern. The solid beam efficiency measured the ability of the antenna to respond to its intended pixel coverage and not the surrounding radiometric background. High solid beam efficiency values require both high polarization purity and attention to reducing the antenna sidelobes by tapering the aperture’s amplitude distribution and maintaining a uniform phase distribution.

A second type of radiometric is referred to as a *sounder*, which uses simple broad coverage antennas. Sounders use frequencies corresponding to the molecular absorption frequencies. The center frequency corresponds to the peak absorption value and several IF bandwidths are used. As the IF bandwidth increases, the absorption at the band edges decreases. Using different IF bandwidths results in a height profile of the radiometric response results, meaning that wider IF bandwidths respond to absorption characteristics at lower altitude levels.

44.3 USER SEGMENT ANTENNAS

Like the space segment, the user segment of satellite systems has a wide array of antenna technology. The early user segments required relatively large ground terminal antennas because of the limited satellite performance, the broad coverage space segment antennas necessitated by uncertain satellite attitudes, and the high receiver noise temperatures of early designs. The large ground terminals generally served as a hub for distribution by terrestrial segments. Improvements in satellite capabilities and user technology availability results in satellite services presently being extended to the personal level with examples being cost effective handheld GPS receivers and direct broadcast services. Such trends can be anticipated to continue in the future. These trends prompt the development of cost effective designs to control total system costs and appeal to individual users. These trends also pose interesting challenges in testing on a production level.

At a system level, like the space segment antennas, the user segment antennas have G/T and ERP requirements to meet the specified link performance. Unlike the space segment antennas, which are required to meet G/T and ERP performance over a specified coverage area, the user segments must align their main beams with the satellite, so that the peak antenna gain values burdened with the antenna pointing loss determines the link performance.

While the user segment exploits a diversity of antenna technology, common issues exist. These issues include antenna temperature characterization, interference protection, duplexing, and production issues.

Antenna Temperature

Unlike space segment antennas, the antenna temperature for the user segment depends on the propagation path and its loss. Propagation loss, in common with ohmic loss in RF circuitry and antenna feeds, both attenuates the received signal and generates additional noise. In both cases, the increased noise is a consequence of blackbody radiation. At microwave frequencies, propagation loss depends on the water vapor and oxygen absorption characteristics in the atmosphere.³⁴ At higher EHF frequencies, the propagation loss is weather dependent.³⁵⁻³⁶ The principal impact results from raindrops within the propagation path, because their size becomes comparable to the EHF wavelength. The loss results from the absorption cross section of rain drops. Other hydrometeors, such as fog, are smaller, and then there are ice crystals and snow, which are both even smaller and have reduced absorption. Statistical projections of path loss at EHF frequencies are commonly made using climate models of the geographical locations based on cumulative rainfall rate statistics. The specific attenuation (dB/km) of the path has an aR^b dependence,³⁷ where R is the rain rate (mm/hr), and the a and b constants depend on the operating frequency. Using the cumulative rain rate statistics and a freezing height, above which rain drops become ice crystals with little loss, projections can be made of link availability. In addition to loss in the propagation path, antennas enclosed by radomes also suffer performance degradation when radome surfaces are wet³⁸⁻³⁹ and must be factored into projections of link availability. These link availability statistics are stationary only in a long-term multi-year sense—for example, some years are drier than others. While such analyses provide a basis of establishing rain margin requirements, the availability of a given system at a particular time is subject to meteorological conditions prevailing at that time.

Like other loss components, such as ohmic loss, the propagation loss not only reduces the signal level but also increases the system noise level. The increase in the system noise temperature for user antenna contrasts with uplink satellite receiving antennas where the 290 K earth background antenna is not increased by noise contributions from loss. Practical antenna systems likewise contain ohmic loss and other necessary components such as filters

unavoidably contain loss. Today's availability of low noise receiver technology often results in receiver noise temperatures that are less than the total antenna noise temperature. Because the user's G/T dictates the received signal sensitivity, attention must be paid to the total system noise temperature T that equals the sum of the antenna temperature and the receiver temperature. In the past, the achievable receiver noise temperature greatly exceeded the antenna noise temperature and consequently, simply reducing the receiver noise temperature almost directly reduced the system noise temperature. This is not the case with available low noise receiver technology.

Antenna temperature for user systems depends not only on the operating frequency but also the propagation path. The loss increases as the elevation angle to the satellite decreases, because the path traveling through atmosphere where loss occurs is longer. The antenna temperature depends on the emission temperature of the environment surrounding the antenna. Finally, practical antennas are not ideally lossless and matched to the receiver input impedance, and corrections must be made for those effects.

The total system noise temperature must be referenced to a specified elevation angle because the antenna temperature depends on the path length through the atmospheric loss. At low elevation angles, the antennas are limited by obscura from natural terrain and man-made structures specific to the site. As a consequence, a moderate elevation angle, for example, 20° , is often selected for specification purposes. Such a choice allows compliance measurements without the measurement uncertainties at the lower elevation angles. In practice, however, the elevation angles used operationally must be addressed and at the lower elevation angles, the antenna temperature can depend on the azimuth angle because of obscura; in such cases, blockage may be a further problem. These concerns must be addressed when a site selection is made for user antennas.

The antenna and receiver noise temperatures comprise the total system noise temperature. The user terminal's G/T requires determining the antenna gain and the system noise temperature at the same terminal reference plane. In many cases, the input to the receiver's LNA is a convenient terminal plane because the receiver temperature can be measured at the LNA input, using standard Y factor hot/cold load measurements. Additionally in many cases, the LNA input is convenient to terminate in a matched load to provide a stable noise temperature value for measuring the antenna temperature.

The total system temperature for a terminal is comprised of the antenna temperature at the reference plane and the receiver noise temperature at the reference plane. The system temperature at this reference plane⁴⁰ T_s is given by

$$T_s = (1 - \Gamma^2) [T_{\text{ant}}L + 290(1 - L)] + T_{\text{rec}}$$

where Γ is the magnitude of the reflection coefficient (for well matched systems, $(1 - \Gamma^2)$ is very close to 1), T_{ant} is the antenna temperature, L is the loss, and T_{rec} is the receiver noise temperature. Physically, the noise power received by a lossless antenna is reduced by that loss but the noise generated by the loss must be included. If significant impedance mismatch loss $(1 - \Gamma^2)$ exists, the noise incident on the LNA must be reduced by the mismatch loss.

The antenna temperature can be established through measurement. If such measurements are not available, the antenna temperature of a lossless antenna can be calculated from the pattern characteristics and the emission background levels. Such a procedure has been done⁴¹ and verified by measurement of a Cassegrain reflector antenna, as shown in Figure 44-9. The antenna temperature differs from the sky temperature values,³⁴ particularly at the lower elevation (high zenith) angles, because of ground emission coupling through the antenna sidelobes. The sky temperature values correspond to the noise received by an ideally lossless antenna with an infinitely small beamwidth. The measured and calcu-

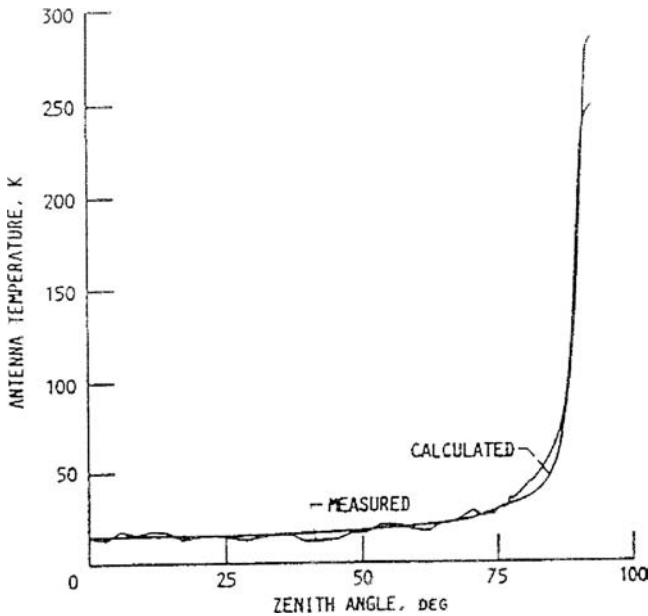


FIGURE 44-9 Calculated and measured antenna temperature (after K. M. Lambert and R. C. Rudduck⁴¹© *Radio Science* 1992)

lated antenna temperatures were performed using very low loss antenna. Practical antenna systems must be corrected for loss in filters and other RF components needed for system operation. Often, data for only the antenna is available and when determining the G/T the gain must be reduced by the RF loss and the antenna temperature likewise compensated by the RF loss as indicated.

Interference Issues

The utilization of satellite and terrestrial systems has greatly expanded in recent years, a trend that can be expected to continue. As a consequence, interference between systems increases both as a possibility and as a concern. Two distinct types of interference problems exist. One type results from interference between satellite programs and the second type is interference to/from user terminals and other terrestrial systems.

The dense population of satellites in geostationary orbits prompted concerns of interference between users and satellites, for example, user illumination of unintended satellites create interference at the unintended satellite and likewise, user reception, from not only the desired satellite but also unintended satellites, creates interference. Accordingly, constraints were placed on the sidelobe envelope for user antennas, and are commonly followed by antenna vendors. A typical envelope requirement⁴² is

$$\begin{aligned}
 G &= 32 - 25 \log \theta, \text{ dBi}, 1^\circ < \theta < 48^\circ \\
 &= -10 \text{ dBi}, 48^\circ < \theta < 180^\circ \\
 &\text{for } D/\lambda > 100
 \end{aligned}$$

and

$$\begin{aligned}
 G &= 52 - 10 \log (D/\lambda) - 25 \log \theta, \text{ dBi}, (100D/\lambda)^\circ < \theta < 48^\circ \\
 &= 10 - 10 \log (D/\lambda), \text{ dBi}, 48^\circ < \theta < 180^\circ \\
 &\text{for } D/\lambda < 100
 \end{aligned}$$

where D/λ is the antenna aperture size in wavelengths and θ is measured from the antenna's boresight axis.

The spectral crowding at the lower microwave frequencies results in potential interference between user satellite antennas and other terrestrial antennas. While frequency regulations are imposed to isolate users, the proximity of terrestrial systems to users is much closer than satellites. While emissions from terrestrial users are isolated in frequency, the higher power levels from nearby terrestrial systems can degrade user systems by saturating the receiver at out-of-band frequencies. Such interference is unintentional. Military systems also have concerns of intentional interference or jamming, and have developed effective means for negating such interference.

User systems can be protected from interference by a variety of techniques. In many cases, site alternatives can be effective in avoiding interference; and spectrum analyzer instrumentation⁴³ is commonly used to determine potential interference sources at a given site. Another operational alternative that can be effective is determining the feasibility of restrictions on the operational use of the interference source (sector blanking for nearby radars, for example), not transmitting in a limited angular sector, is sometimes effective in controlling interference from radar systems.

In addition to such operational alternatives, four basic techniques may be applied to user systems to negate interference. The gain and filtering distribution of the receiver should be examined to filter interference so that the receiver remains in its linear operating range is a basic technique that should be addressed in the terminal design. Spread spectrum modulation and data interleaving are commonly used to dilute and distribute the effects of interference.

In addition to the receiver architecture and the modulation alternatives, two antenna techniques are also available to negate interference. One technique is the passive control of antenna sidelobes.⁴⁴⁻⁴⁶ In many cases, the interference arrives through the antenna sidelobes. Because the radiation mechanisms for the wide angle sidelobes do not contribute to the aperture distribution, these sidelobe control techniques can be employed without significantly impacting the antenna's main beam gain performance. A second technique is the adaptive cancellation of interference. For user antennas, adaptive sidelobe cancellation⁴⁷⁻⁴⁸ is an appropriate design choice. In this design, auxiliary antenna elements with gain levels that exceed the sidelobe gain are used to sample the interference power. Correlation of the auxiliary antenna elements with the main antenna response is used as a means to dynamically determine the adaptive weight settings. Since the main antenna output is made up of both desired signal and interference components, whereas the auxiliary antenna elements have predominantly interference. The control circuitry for the adaptive weight values is based on minimizing the correlation of the main antenna and the auxiliary antenna elements. The gain of the auxiliary antenna elements and the circuitry is purposely constrained to avoid canceling the main beam of the main antenna containing the desired signal power. One problem with such designs is that the frequency response of the main antenna does not match the response of the auxiliary antenna elements nor are their phase centers coincident. Effective interference over a bandwidth requires equalization of the responses of the main and auxiliary antennas. The differences in these antenna responses have been examined⁴⁹ and generally, an adaptive transversal equalizer is used to equalize these antenna responses. The number of taps and the required tap spacings for the adaptive transversal filter design have been derived for a specified interference cancellation performance.

44.4 SATELLITE ANTENNA TESTING

Spacecraft antennas require rigorous RF and environmental testing because of the reliability needed to perform over their long operational lifetimes. While antenna measurement techniques are well established,⁵⁰ spacecraft antennas have some specialized test requirements that are not commonly used in other applications. User segment antennas also pose antenna test challenges, particularly testing in a production basis as services are extended to a personal level. Larger ground terminals also present test challenges, particularly from excessive far field requirements. Radio source techniques are commonly used in such situations.

Spacecraft Antenna Test Phases

There are three distinct phases in testing spacecraft antennas, as illustrated in Figure 44-10. The first phase is developmental testing, where the compliance of candidate design(s) with system requirements are established; requirements include both RF performance and testing for compatibility with the launch and on-orbit environments. The second phase is qualification testing, which verifies that the flight hardware conforms to the RF performance established during development testing and that the flight hardware can withstand launch and on-orbit environments. The third phase is on-orbit testing that has the objectives of establishing compliance with system specifications, providing baseline data for subsequent trending, and complementing satellite telemetry to diagnose on-orbit performance shortcomings that may arise during the satellite's lifetime.

Each test phase differs in both scope and requirements. Additionally, the test parameters in each of these phases evolve from component level parameters in the development phase to system level parameters used in the on-orbit phase. Detailed testing can be conducted using general purpose instrumentation and facilities in the development phase, while specialized ground terminals that are generally part of the program's mission control terminal are required for on-orbit testing and the satellite in its orbital location rather than in a test facility.

Development Testing

Development testing can generally be accomplished with general purpose test facilities and instrumentation, and it uses engineering model hardware. Much of the RF testing is

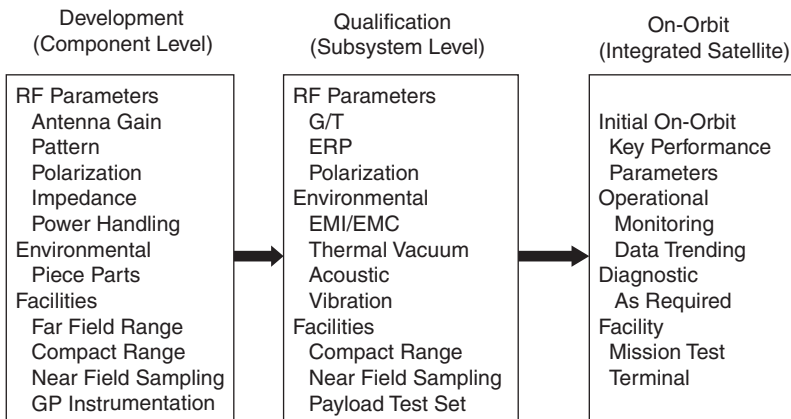


FIGURE 44-10 Satellite test phases

conducted at the component level and detailed measurements on the engineering model are performed. Typically, the projected performance of the antenna is evaluated using analytic codes; agreement between these analytic projections and engineering model measurements provides confidence in the results. In some cases, more than one candidate design exists at the initiation of the program, and design tradeoffs are conducted during the development phase to select a final design. In other cases, requirements for other issues such as electromagnetic interference/electromagnetic compatibility (EMI/EMC) result in test requirements at out-of-band frequencies or in specialized models needed to emulate the spacecraft structure and/or payload antennas. Examination of the candidate designs for identifying potential workmanship issues is a prudent practice for assisting in flight hardware testing. Finally, present testing is often conducted using software test procedures and experience with these procedures provides benefits for subsequent qualification testing.

At the initiation of the program, a requirements verification matrix is generally constructed flowing down the specific antenna requirements from the system specification and identifying how these requirements are to be verified at each phase of the testing. Attention to the test requirements identified in this matrix is necessary early in the program's initiation in order to identify instrumentation and facility needs, as well as necessary test points within the antenna hardware. In some cases, development is required to conduct the qualification testing. Additionally, the requirements verification matrix helps to identify the scope of the test procedures and planning. Such steps are necessary in determining program needs and schedule requirements.

Qualification Testing

Flight hardware must be space qualified to demonstrate its capability for properly performing over its orbital lifetime. Qualification testing is performed on flight hardware that must be properly protected. RF testing is therefore performed in protected test facilities to avoid damage to flight hardware. Compact ranges and near field sampling facilities are in common use for this testing. Qualification testing also requires demonstrating that the flight hardware is capable of withstanding launch and on-orbit environments. These environments typically require acoustic, thermal and vacuum, and vibration testing. Such testing is often performed at a more stringent level, on engineering development hardware that emulates the flight hardware in order to determine design margins. The flight hardware is then tested at a somewhat lower qualification level following environmental values anticipated during launch and on-orbit operations. One of the important features of the qualification level testing is the identification of workmanship errors (such as cable integrity) and specific payload assembly errors that would limit on-orbit reliability. Such testing benefits from identification of potential workmanship errors in the development test phase.

Qualification testing is generally conducted at the component, subsystem, and integrated spacecraft levels. The RF testing is distributed over these levels with attention at the component and subsystem levels using common antenna measurement parameters and at the integrated spacecraft level, transitioning to system level parameters such as ERP, G/T , and BER values for the overall transponder. Redundant electronics paths are also evaluated in this testing, but this involves not just the performance of these redundant paths but the command paths needed to select the redundant components as well. Testing of this type particularly within the test environments precludes the normal RF test facilities. Testing in these environments often involves "hat couplers" that are comprised of RF sealed enclosures containing RF probes to allow test signal injection and reception. Such hat couplers permit testing within environmental facilities such as thermal vacuum chambers that preclude conventional test facilities. The hat coupler designs are specific to the antenna technology used, and require development and calibration prior to their use in

qualification testing. Likewise, testing to evaluate both the antennas and payload electronics is typically desired after shipment to launch sites and spacecraft assembly areas. The development of suitably portable RF test facilities would benefit testing in these areas, and as the spacecraft antennas become more complex and integrated with RF electronics, test requirements for the assembled payload will evolve from a “light bulb” evaluation of the health to more insightful evaluations of the antenna systems functionality.

The evaluation of EMI/EMC compliance is a spacecraft-wide issue. At the initiation of the spacecraft design, the frequency plans should be examined to avoid conflicts. All frequency components and their harmonics need to be examined and the coupling between these components and other potentially susceptible components throughout the entire spacecraft need to be examined. Estimates of these potential coupling paths need to be examined and those paths with limited susceptibility margins need to be identified and verified by test. It is not unusual for the number of potential coupling paths to exceed 10,000, and after examination the number of paths with limited margin will be substantially less. The spacecraft antennas are generally well isolated from the majority of the payload electronics and their coupling can be estimated and/or augmented by measurement particularly at out-of-band frequencies that are not normally characterized. Some of the coupling paths can be compromised by workmanship errors and call for examination of the flight hardware. Some EMI/EMC testing is performed at the component and subsystem levels. Generally, the assembled flight hardware is tested in an EMI/EMC facility to assure compliance. The natural background levels within these facilities needs to be established first, and these levels compared with the required sensitivities. Emissions from the test equipment and any anticipated leakage of the cabling that supports the testing bears scrutiny prior to the spacecraft examination. The EMI/EMC testing should concentrate on those paths previously identified as well as a more general survey of the emissions. Measurements of potential emissions are often sniffed using waveguide probes at microwave frequencies to detect leakage components that could degrade system operation.

Satellite qualification testing includes some specialized evaluations that are not normally a concern in other antenna applications. One concern for spacecraft transmitting antennas is multipaction⁵¹ that results in vacuum conditions when sufficient electrical field strength exists to strip electrons from metallic surfaces. A resonant condition produces an electron avalanche damaging the surfaces. Multipaction susceptibility depends on the product of gap size between surfaces and the RF frequency the spacing between surfaces, vacuum conditions, and the secondary electron emission properties of the materials as well as the RF power levels. Multipaction generally occurs at the smallest separation between surfaces in components. Testing components with potential multipaction concerns is generally performed under vacuum conditions, and since a 6 dB margin in power handling capability is commonly used, the transmitter power for such testing is accordingly higher than that used operationally.

Space segment antennas serve a variety of users and thus typically have modulations with multiple carrier components. In addition to transmitter linearity requirements, concerns also exist regarding the linearity of the antennas and RF components, particularly in the case where the same antenna aperture is used to both receive and transmit; the receive and transmit bandwidths are isolated by the filtering in a diplexer. Mechanical joints within the RF and antenna components can exhibit nonlinear behavior creating intermodulation products that interfere with signal reception. This phenomenon is referred to as (passive intermodulation) PIM.⁵²⁻⁵³ Electroforming techniques can minimize the required number of joints in RF components and reduce potential PIM levels. The generation of PIM products is often determined by injecting two CW tones into the components. The frequencies are chosen to produce PIM products within the received passband. An examination of various orders of potential PIM products is typically performed to understand what problems can arise and guide the evaluation of potential PIM generation.

Another concern for satellite antennas is the potential of electrostatic discharge (ESD), which can damage or interfere with RF electronics. Satellite antennas are generally protected from the on-orbit thermal extremes by thermal covers. Such covers necessarily must have low insertion loss and are typically dielectric materials. On-orbit, such covers can accumulate sufficient charge to produce a discharge⁵⁴⁻⁵⁵ and while the spectra of the discharge has strong low frequency components, sufficient energy can couple into RF electronics at levels that can cause damage or interference. In other cases, the same thermal control materials are used to protect other spacecraft components such as solar arrays, and while the discharge is not associated with the antenna itself, nonetheless, spectral coupling from the discharge into the antenna must be considered to evaluate potential interference. The potential for ESD is particularly high during geomagnetic storms. Thermal covers, therefore, require sufficient conductivity in either their composition or through special ESD paints and ground paths to control the charge buildup. Generally, ESD characteristics are measured on material samples that are charged with electron beams to determine the spectral characteristics of the discharge. Such tests are often performed on samples of the materials and in some cases, on the assembled antenna to verify the integrity of the grounding paths.

A common problem involves testing antennas that have broad coverage requirements such as the hemispheric coverage required by TT&C antennas. Because these designs often use lower microwave frequencies and must be compact, isolating such antennas from the surrounding spacecraft becomes a challenge. The antenna performance on the spacecraft differs from its free space performance and the interaction between the antenna and spacecraft results in pattern ripples in the coverage area that can lead to noncompliance. Antenna backlobe reduction helps to reduce the interaction. Testing on the spacecraft has its challenges not only from the physical size of the spacecraft but also the increased far field distance requirements compared to antenna itself. Development attention to address analysis and measurement techniques for such situations is recommended.

The environmental testing on spacecraft components includes acoustic, vibration, and thermal vacuum evaluations. Such testing is performed on the antennas by themselves and repeated when the antennas are integrated with the payload. In cases where the design margins relative to the environmental tests are required, testing performed on engineering model hardware is often performed at more stringent protoqualification levels. The test levels depend on the launch vehicle characteristics and structural details of the payload and the thermal characteristics projected in orbit. These tests are useful in ascertaining workmanship defects such as cable shortfalls. Testing within the environmental facility, particularly for the assembled payload, often require an end-to-end evaluation, such as BER measurements. Since conventional antenna test facilities cannot be configured within the environmental facility, a means of injecting test signals and measuring their output is required. In some cases, test couplers are integrated into the antenna hardware for such purposes and in other cases, hat coupler are used to inject or receive test signals from the antenna. Both injection methods require attention to calibration and measurement repeatability to obtain meaningful results particularly for evaluating the noise characterization for uplink receiving antennas.

On-Orbit Testing

On-orbit testing verifies compliance with the key system level parameters for the spacecraft initially, and provides performance monitoring capabilities of the satellite over its lifetime. These measurements are at the system level, such as ERP, G/T , BER, and others. Such testing is performed with a specialized ground terminal. The requirements for these terminals include rigorous calibration and diagnostic capabilities to ensure accurate results, and other capabilities that are specific to the programs requirements. For example, systems that use

polarization reuse place stringent polarization purity requirements on the ground terminal to accurately measure the satellite's polarization isolation. Likewise, some satellite systems have requirements for antenna sidelobe control (such as isolation characteristics) between multiple beams, and the sensitivity required to verify such sidelobe requirements dictates the ground terminal aperture size to perform accurate measurements. Other monitoring requirements include the satellite user loading characteristics versus the satellite transmitter operating point to control the level of intermodulation products.

On-orbit testing requires a dedicated ground terminal to monitor the on-orbit satellite operation. Such terminals are generally a part of the program's mission control facilities. These test terminals require particular attention to calibration to obtain accurate measurements. Additionally, these terminals require effective diagnostic capabilities to assure indicated shortfalls result from the spacecraft rather than the test terminal. Generally, the performance requirements for the test terminal exceed those of the user terminals. For example, measurement of the isolation between antenna beams requires a significantly larger test terminal aperture than the data reception requirements for a user terminal.

User Antenna Testing

User antennas for satellite systems also have requirements that differ from other applications. One problem that will become more commonplace is the ability to test at production levels. This problem results as satellite services are extended more broadly at the personal level and the number of user terminals increases. Testing each individual antenna is often both cost and schedule prohibitive. One approach for testing at a production level examines those areas where workmanship shortfalls can lead to noncompliance. One example of such shortfalls is user terminals where polarization reuse is used and the polarization purity must be maintained to isolate signals on orthogonal polarizations. An example of workmanship shortfall leading to noncompliance of these isolation requirements is the failure to maintain the required mechanical tolerance in the internal antenna feed dimensions that can increase cross-polarization levels. In this case, simple mechanical tolerance testing and random RF testing can provide assurance of compliance with polarization purity requirements without testing each individual feed.

Compliance testing for user antennas is generally conducted using general purpose instrumentation and test facilities. The normal antenna characterizations of gain, polarization, bandwidth, and impedance parameters are conducted using standard measurement practices. Other requirements, such as sidelobe envelope compliance, must address the capabilities needed for the sensitivity for accurately measuring lower power levels from the sidelobes and attention to the facility multipath levels that can degrade measurement accuracies. In such cases, attention to leakage levels in the RF test instrumentation and microwave components are needed to assure the apparent sidelobe levels are from the antenna rather than leakage components. The correlation of these sidelobe responses with analytic projections of the design performance provides increased measurement credibility.

Radio Source Measurements

A measurement technique using radio sources has proved useful for characterizing the G/T of user antennas. Originally, the technique was applied to large ground terminal antennas,⁵⁶⁻⁵⁷ which have a physical size that cannot be accommodated in general purpose test facilities and which have far field distances that are excessive. Because of available low noise receivers in system use, the technique can be applied to much smaller diameter antennas⁵⁸ and, as RF electronics are integrated into the antenna design, it can provide a means of testing designs that have no convenient antenna terminals.

The measurement technique⁵⁹ consists of measuring the noise power from a radio source and measuring the cold sky background noise at the same elevation angle; the ratio of these noise power measurements is a Y factor. The noise flux densities of radio sources are very well known. For large antennas, radio sources such as Cassiopeia A are used. If the antenna does not have the sensitivity to measure the noise from these sources, the moon is commonly used. The strongest radio source is the sun, and measurements can be made on much smaller antennas using this source. The sun has some limitations. Solar flare activity results in variations in the noise flux densities, but measurements of the solar flux are made daily by a global network, so that variations in the solar flux are determined. The solar disk subtends about a $1/2^\circ$ angular width so that it does not appear to be a point source for antennas whose beamwidths approach the $1/2^\circ$ width. Techniques to correct for the angular width have been devised. However, because of solar flares, the solar flux density is not uniformly distributed over the solar disk, and experience has shown that particularly during active solar periods, measurements of antennas with beamwidths smaller than $1/2^\circ$ experience experimental results that are not always repeatable. By contrast, the correction for the angular width of the moon is not limited because the lunar flux results from solar reflection and does not have the variability of flare activity. The G/T from these measurements is derived from

$$G/T = 8\pi k (Y - 1)/S\lambda^2$$

where k is Boltzmann's constant, Y is the noise power ratio of antenna pointed at the source and the antenna pointed at a cold sky background, S is the radio source flux density, and λ is the wavelength.

In practice, the G/T is referenced to a specified elevation angle because of the antenna noise temperature dependence on elevation angle. Since the radio source is generally at a different elevation angle, the G/T must be compensated for the differences in the cold sky noise measurement where the source is located, as well as the cold sky noise measurement at the specified elevation angle. The radio source flux density does not include the propagation loss of the atmospheric path and compensation for this path loss must be made. At microwave frequencies, the path loss can be estimated, but at the higher EHF frequencies, the path loss depends on weather conditions. A separate measurement of the antenna temperature versus elevation angle is made as a means of determining weather effects at the time of the G/T measurement. The antenna temperature can be measured from the ratio of the noise powers when the receiver is terminated by an ambient load and when the receiver is terminated by the antenna pointed at the desired elevation angle. This ratio equals the sum of the ambient and receiver noise temperatures divided by the sum of the antenna and receiver noise temperatures. The antenna temperature is determined from this ratio and the receiver noise temperature. The G/T measurements need to assure the antenna is properly pointed at the noise source and that the measurements are not degraded by interference.

In many cases, the antenna gain rather than the G/T is required. The total system temperature is the sum of the antenna noise temperature and the receiver noise temperature. The antenna temperature can be measured and the receiver noise temperature is typically measured using the standard Y factor technique with hot and cold loads. The antenna gain is determined from the product of the measured G/T and the system noise temperature comprised of the sum of the antenna temperature at the elevation angle where the radio source was located for the G/T measurement and the receiver noise temperature. The sensitivity of the antenna temperature to elevation angle is of interest not only to obtain the correct value for determining the antenna gain but also in understanding how the user antenna system will perform at different elevation angles. The antenna gain can then be determined by multiplying the antenna's G/T by the system noise temperature. In addition, a common reference plane must be used for both the antenna's G/T and system noise temperature values. In some cases, the antenna gain within the transmit band is desired. If an LNA

operating in the transmit band is substituted for the transmitter, a G/T measurement using a radio source can be performed. Determining the system noise temperature T for this case allows the determination of the transmit antenna gain.

Diplexer

User antennas commonly require the same antenna for both receive and transmit operation. A diplexer is required to separate the transmitting and receiving frequencies. While a variety of techniques may be used for diplexing, the requirements for diplexing are common to all designs. A diplexer provides the filtering necessary to isolate the receiver and transmitter and a means to connect the receiver and transmitter to the antenna. Diplexer loss is to be minimized to avoid reduced received and transmitted signal levels. The requirements of filtering are derived in the following manner. The transmit filtering must pass the transmitted spectrum with low loss and with sufficient amplitude and phase flatness to avoid transmit signal distortion. The transmit filter must sufficiently suppress the out-of-band emissions from the transmitter that fall within the receive bandwidth sufficiently to avoid degrading receiver performance. This filtering requirement requires knowledge of the out-of-band spectral characteristics of the transmitter and generally, the required filter rejection in the receive bandwidth reduces the out-of-band transmitted spectrum to a level 10 dB below the receiver noise spectra density. This requirement also forms a basis to test the rejection by observing the receiver noise level with the transmitter off and the increase in the receiver noise level with the transmitter on. Testing for high transmit power levels also involves examining the power handling capabilities. Two requirements exist for high transmit power levels and are reflected in the testing. The first requirement is to ensure breakdown does not occur, a peak power limitation. The second requirement concerns average power limitations and involves demonstrating the increased temperature resulting from ohmic loss stays beneath a safe operating temperature limit. In such cases, the transmit diplexing filtering might be distributed with some of the filtering being located on the transmitter and other filtering located near the antenna's diplexer combiner to distribute the thermal load.

The receive filter must pass the receive spectrum with low loss, as well as sufficient amplitude and phase flatness to avoid receive signal distortion. The filtering must also suppress the transmitted spectrum sufficiently to avoid compressing the receiver at out-of-band frequencies. This filtering requirement involves determining the spectral coupling of the transmitter into the receiver port, understanding the gain and filtering distribution of the receiver, and determining the power levels needed to saturate the receiver at out-of-band frequencies. Again, the filtering rejection is sized with a margin, commonly 6 dB, to avoid receiver compression throughout the receiver architecture. Testing involves measuring the receiver's amplifiers output levels at the out-of-band frequencies where saturation could potentially occur.

In practice, a significant amount of filter rejection can be required. Attention should be given to leakage around filter sections to assure the overall filtering rejection is not compromised. Leakage components can be detected by using waveguide probes to assure component joints and tuning screws do not have excessive leakage levels. Electroforming techniques and distributing the filters at different locations can be helpful. The overall filter rejection may be sufficiently high to limit verification within the dynamic range of standard instrumentation such as network analyzers. The network analyzers provide the means to measure in-band insertion loss and amplitude and phase flatness as necessary but determining the rejection can require further sensitivity. Using additional amplification may provide the added sensitivity and in the case of distributed filtering, the individual components can be measured. In any event, measurements of the assembled system are necessary to assure the diplexer performance complies with requirements.

Radiation Hazard

Large ground terminal antennas often require significant power levels to meet link requirements for high data rate transfer. When high transmitting power levels, concerns exist that the power densities near the antenna can potentially harm personnel. Generally, the safe levels for RF exposure⁵⁰ are 2 mW/cm² and the debate on the safe level is ongoing. When such concerns exist, measurements of the antenna fields are performed. These measurements are typically within the antenna's near field and are performed using a probe antenna, as well as spectrum analyzer instrumentation. Because the power densities of concern have extremely high levels, measurements are generally performed using a signal generator in place of the actual transmitter to avoid concerns of potential harm to personnel performing the measurements and generally adequate sensitivity exists when the signal generator is used.

Such measurements are guided by first using available analysis codes to examine the near field characteristics to identify regions where the near field power densities could potentially exceed the safe levels of exposure. One region is the near field of the aperture distribution and since in operation, the antenna is pointed at the satellite, this region is not readily accessible to personnel on the ground. Other regions, for example, pillover and scattering by struts supporting the feed and/or subreflector, are potential near field locations where the safe levels can be exceeded. Using the near field analyses provides insight of where measurements should be performed and an indication of what their levels should be. Other regions need to be examined if their effects are not included in the analyses. For example, leakage through the reflector panel gaps in large reflector antennas and openings in the reflector surfaces for antennas feeds need to be separately probed.

REFERENCES

1. D. H. Martin, *Communication Satellites-Fifth Edition* (El Segundo, CA: The Aerospace Press and AIAA, 2007).
2. E. G. Njoku and E. K. Smith, "Microwave Antenna Temperature of the Earth From Geostationary Orbit, *Radio Science*, vol. 20 (May-June 1985): 591-599.
3. C. T. Brumbaugh, A. W. Love, G. M. Randall, D. K. Wainee, and S. H. Wong, "Shaped Beam Antenna for the Global Positioning Satellites," *1976 IEEE AP-S Symposium Digest* (October 1976): 117-120.
4. T. Katagi and Y. Takeichi, "Shaped Beam Horn Reflector Antenna," *IEEE Trans Antennas and Propagation*, vol. AP-23 (November 1975): 757-763.
5. A. R. Cherrette, S. W. Lee, and R. J. Acosta, "A Method for Producing a Shaped Contour Radiation Pattern Using a Single Shaped Reflector and a Single Feed," *IEEE Trans Antennas and Propagation*, vol. 37 (June 1989): 698-706.
6. O. M. Bucci, G. D'Elia, G. Mazzarello, and G. Panariello, "Antenna Pattern Synthesis: A New General Approach," *Proc IEEE*, vol. 82 (March 1994): 358-371.
7. R. B. Dybdal, "Polarization Efficiency Statistics," *1999 IEEE MILCOM Symposium Digest* (November 1999): 203-207.
8. R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays* (New York: Wiley, 1980).
9. K. M. Soo Hoo and R. B. Dybdal, "Resolution Performance of an Adaptive Multiple Beam Antenna," *IEEE MILCOM '89 Symposium Digest* (October 1989).
10. J. T. Mayhan, "Area Coverage Adaptive Nulling from Geosynchronous Satellites: Phased Arrays Versus Multiple Beam Antennas, *IEEE Trans Antennas and Propagation*, vol. AP-27 (March 1986): 410-419.
11. R. B. Dybdal, D. J. Hinshilwood, and K. M. SooHoo, "Development Considerations in the Design and Simulation of Adaptive MBAs for Satellite Communications," *1993 IEEE MILCOM Symposium Digest* (October 1993).

12. R. B. Dybdal and H. J. Wintroub, "An Antenna Based High Data Rate Concept," *IEEE MILCOM '95 Symposium Digest* (November 1995).
13. R. B. Dybdal and S. J. Curry, "Adaptive Beam Pointing," *1996 IEEE MILCOM Symposium Digest* (October 1996): 882–886; see also, R. B. Dybdal and S. J. Curry, Adaptive Receiving Antenna for Beam Repositioning, (April, 14, 1998): U.S. Patent 5,739,788.
14. K. Ueno, T. Itanami, H. Kumazawa, and I. Ohtomo, "Design and Characteristics of a Multiband Communication Satellite Antenna System," *IEEE Trans Aerospace and Electronic Systems*, vol. 31, 600–606
15. R. J. Mailloux, *Phased Array Antenna Handbook*, Chap. 8 (Norwood, Massachusetts: Artech, 2005).
16. P. Ingerson and C. A. Chen, "The Use of Non-Focusing Aperture for Multibeam Antenna," *1983 IEEE AP-S Symposium Digest* (May 1983): 330–333.
17. K. S. Rao, G. A. Morin, M. Q. Tang, S. Richard, and K. K. Chan, "Development of a 45 GHz Multiple Beam Antenna for Military Satellite Communications," *IEEE Trans Antennas and Propagation*, vol. 43 (October 1995): 1036–1047.
18. A. W. Love (ed.), *Reflector Antennas* (New York: IEEE Press, 1978).
19. R. B. Dybdal, S. J. Curry, and M. A. King, "An Uplink Multiple Beam Concept for Theater Coverage," *2002 IEEE MILCOM Symposium Digest* (October 2002).
20. R. B. Dybdal and S. J. Curry, "An Uplink Antenna for Electronic Beam Steering and Interference Reduction," *2002 IEEE AP-S Symposium Digest*, San Antonio TX, vol. 1 (June 2002): 590–593.
21. R. B. Dybdal, "Adaptive Control of Multiple Beam Satellite Transponders," *1997 IEEE MILCOM Symposium Digest* (November 1997): 252–255; see also, R. B. Dybdal, Adaptive Control of Multiple Beam Communication Transponders, (April 25, 2000): U. S. Patent 6,055,431.
22. F. J. Dietrich, P. Metzen, and P. Monte, "The Globalstar Cellular Satellite System," *IEEE Trans Antennas and Propagation*, vol. 46 (June 1998): 933–942.
23. J. J. Schuss, J. Upton, B. Myers, T. Sikina, A. Rohwer, P. Makridakas, R. Francois, L. Wardle, and R. Smith, "The IRIDIUM Main Mission Antenna Concept," *IEEE Trans Antennas and Propagation*, vol. 47 (March 1999): 416–424.
24. T. J. Peters, "A Conjugate Gradient-based Algorithm to Minimize the Sidelobe Level of Planar Arrays with Element Failures," *IEEE Trans Antennas and Propagation*, vol. 39 (October 1991): 1497–1504.
25. R. Miura, T. Tanaka, I. Chiba, A. Horie, and Y. Karasawa, "Beamforming Experiment with a DBF Multibeam Antenna in a Mobile Satellite Environment," *IEEE Trans Antennas and Propagation*, vol. 45 (April 1997): 707–714.
26. See Chap. 42.
27. H. J. Liebe, "Modeling Attenuation and Phase of Radio Waves in Air at Frequencies Below 1000 GHz," *Radio Science*, vol. 16 (November–December 1981): 1183–1199.
28. H. J. Liebe, "An Updated Model for Millimeter Wave Propagation in Moist Air," *Radio Science*, vol. 20 (September–October 1985).
29. R. B. Dybdal and F. I. Shimabukuro, "Electronic Vulnerability of 60 GHz Crosslinks," *1984 IEEE MILCOM Symposium Digest*, Paper 27.6 (October 1984).
30. D. E. Ping, J. T. Shaffer, L. U. Brown, and R. B. Dybdal, "A Broadband Rolled Edged Cavity Antenna," *2004 IEEE AP-S Symposium Digest* (June 2004).
31. R. Tomiyasu, "Remote Sensing of the Earth by Microwaves," *Proc IEEE*, vol. 62 (January 1974): 86–92.
32. ..., Special Issue of SIR-C/X SAR, *IEEE Trans on Geoscience and Remote Sensing*, vol. 33 (July 1995).
33. ..., Special Issue on the Windsat Spaceborne Polarimetric Radiometer, *IEEE Trans Geoscience and Remote Sensing*, vol. 44 (March 2006).
34. E. K. Smith, "Centimeter and Millimeter Wave Attenuation and Brightness Temperature Due to Atmospheric Oxygen and Water Vapor," *Radio Science*, vol. 17 (November–December 1982): 1455–1464.

35. R. K. Crane, *Electromagnetic Wave Propagation Through Rain* (New York: Wiley, 1996).
36. L. J. Ippolito, "Propagation Effects Handbook for Satellite Systems Design," ITT Industries Doc. AF01-0006 (September 2000).
37. R. L. Olsen, D. V. Rogers, and D. B. Hodge, "The aR^b Relation in the Calculation of Rain Attenuation," *IEEE Trans Antennas and Propagation*, vol. AP-26 (March 1978): 318–329.
38. F. J. Dietrich and D. B. West, "An Experimental Radome Panel Evaluation," *IEEE Trans Antennas and Propagation*, vol. AP-36 (November 1988): 1566–1570.
39. C. E. Hendrix, J. E. McNalley, and R. A. Monzingo, "Depolarization and Attenuation Effects of Radomes at 20 GHz," *IEEE Trans Antennas and Propagation*, vol. AP-37 (March 1989): 320–328.
40. R. E. Collin and F. J. Zucker, *Antenna Theory*, Chap. 4 (New York: McGraw Hill, 1969).
41. K. M. Lambert and R. C. Rudduck, "Calculation and Verification of Antenna Temperature for Earth-Based Reflector Antennas," *Radio Science*, vol. 27 (January–February 1992): 23–30.
42. CCIR Recommendation 465–563.
43. R. B. Dybdal, G. M. Shaw, and T. T. Mori, "A RFI Measurement System for Field Sites," *1995 AMTA Symposium Digest* (November 1995).
44. R. B. Dybdal, "Millimeter Wave Antenna Technology," *IEEE Trans Selected Areas on Comm*, vol. SAC-1 (September 1983): 633–644.
45. H. P. Coleman, R. M. Brown, and B. D. Wright, "Parabolic Reflector Offset Fed with a Corrugated Horn," *IEEE Trans Antennas and Propagation*, vol. AP-23 (November 1975): 817–819.
46. A. B. Crawford, D. C. Hogg, and L. E. Hunt, "A Horn Reflector Antenna for Space Communications," *BSTJ*, vol. 40, 1095–1116.
47. P. W. Howells, Intermediate Frequency Sidelobe Canceller, (1965): U. S. Patent 3,202,990.
48. K. M. SooHoo and W. Masenten, "Adaptive Sidelobe Canceller Designs for Large Earth Terminals," *1984 IEEE MILCOM Symposium Digest*, paper 40.5, (Classified).
49. R. A. Dell Imagine and K. M. SooHoo, "Adaptive Sidelobe Canceller Designs for Large Earth Terminals," *Proc 1980 Adaptive Antenna Symposium*, RADC Doc. TR-80-378.
50., *IEEE Standard Test Procedures for Antennas* (New York: Wiley-Interscience, 1979).
51. A. D. Woode and J. Petit, "Design Data for the Control of Multipactor Discharge in Spacecraft Microwave and RF Systems," *Microwave Journal* (January 1992): 142–155.
52. J. W. Boyhan, H. F. Lenzing, and C. Koduru, "Satellite Passive Intermodulation: Systems Considerations," *IEEE Trans Aerospace and Electronic Systems*, vol. 32 (July 1996): 1058–1063.
53. J. W. Boyhan, "Ratio of Gaussian PIM to Two Carrier PIM," *IEEE Trans Aerospace and Electronic Systems*, vol. 36 (October 2000): 1336–1342.
54. C. Bowman, A. Bogorad, G. Brucker, S. Seehra, and T. Lloyd, "ITO-Coated RF Transparent Materials for Antenna Sunshields-Space Environment Effects," *IEEE Trans Nuclear Science*, vol. 37 (December 1990): 2134–2137.
55. H. C. Koons and T. S. Chin, "Broadband RF Spectrum for Electrostatic Discharges on Spacecraft," Aerospace Corp Tech Rept, TR-93(3940)-6, (May 1993).
56. J. W. M. Baars, "The Measurement of Large Antennas with Cosmic Radio Sources," *IEEE Trans Antennas and Propagation*, vol. AP-21, 461–474.
57. D. F. Wait, "Precision Measurement of Antenna System Noise Using Radio Stars," *IEEE Trans Instrumentation and Measurement*, vol. IM-32 (March 1983): 110–116.
58. R. B. Dybdal, "G/T Measurement of Small Antennas," *1997 AMTA Symposium Digest* (November 1997): 37–42.
59. R. B. Dybdal, "On G/T Radio Source Measurements," *2000 AMTA Symposium Digest* (October 2000): 187–191.

Chapter 45

Earth Station
Antennas

William A. Imbriale

*Jet Propulsion Laboratory
California Institute of Technology*

CONTENTS

45.1	INTRODUCTION.....	45-2
45.2	SINGLE-BEAM EARTH STATION ANTENNAS.....	45-4
45.3	MULTIPLE-BEAM EARTH STATION ANTENNAS.....	45-13
45.4	ANGLE-TRACKING TECHNIQUES.....	45-16
45.5	POLARIZATION	45-17
45.6	MEASUREMENT TECHNIQUES	45-18

45.1 INTRODUCTION

Almost all Earth station and ground station antennas are used for either telecommunications or science, and frequently the same antenna (as, for example, the 70-m antenna of the National Aeronautics and Space Administration [NASA] Deep Space Network [DSN]¹) is used for both purposes. The terms Earth station antenna and ground station antenna seem to be used interchangeably. However, most often, antennas used for supporting Earth-orbiting satellites are called Earth station antennas and those used primarily for science or deep-space satellites are termed ground station antennas.

This chapter primarily deals with telecommunications, and Chapter 49 covers the science applications. However, much of the material included in this chapter (e.g., analysis, design, and measurement techniques) applies equally well to both applications and will not be repeated in Chapter 49. This chapter also uses some material from previous editions of the book.

Reflector antennas have existed since the days of Heinrich Hertz (1857–1894) and are still one of the best solutions to requirements for cost-effective, high-gain, high-performance antenna systems. Consequently, virtually all ground station antennas are reflector antennas of one type or another.

Requirements for ground station antennas can be grouped into several major categories:

- Electrical or radio-frequency (RF) requirements
- Control-system requirements
- Structural requirements
- Pointing- and tracking-accuracy requirements
- Environmental requirements
- Miscellaneous, such as those concerning radiation hazards, primary-power distribution for de-icing, etc.

Only the electrical or RF requirements will be dealt with in this chapter.

The primary electrical antenna specifications are gain, noise temperature, voltage-standing-wave ratio (VSWR), power rating, receive–transmit group delay, radiation pattern, polarization, axial ratio, isolation, and G/T (antenna gain divided by system noise temperature). All parameters except the radiation pattern are determined by system requirements. For commercial Earth station antennas, the radiation pattern must meet the minimum requirements set by the International Radio Consultative Committee (CCIR) of the International Telecommunications Union (ITU) and/or national regulatory agencies such as the U.S. Federal Communications Commission (FCC). For antennas used for deep-space communications, the primary design requirement is for G/T , and, in general, there are no requirements for a specific radiation pattern.

Earth station antennas operating in the field of international satellite communications must have sidelobe performance as specified by INTELSAT standards or by CCIR Recommendation 580-1 (see Figure 45-1).

The CCIR standard specifies the pattern envelope in terms of allowing 10 percent of the sidelobes to exceed the reference envelope and also permits the envelope to be adjusted for antennas whose aperture is less than 100 wavelengths (100λ). The reference envelope is given by

$$\begin{aligned}
 G &= [49 - 10 \log (D/L) - 25 \log \theta] \text{ dBi} & D \leq 100\lambda \\
 &= (29 - 25 \log \theta) \text{ dBi} & D > 100\lambda
 \end{aligned}$$

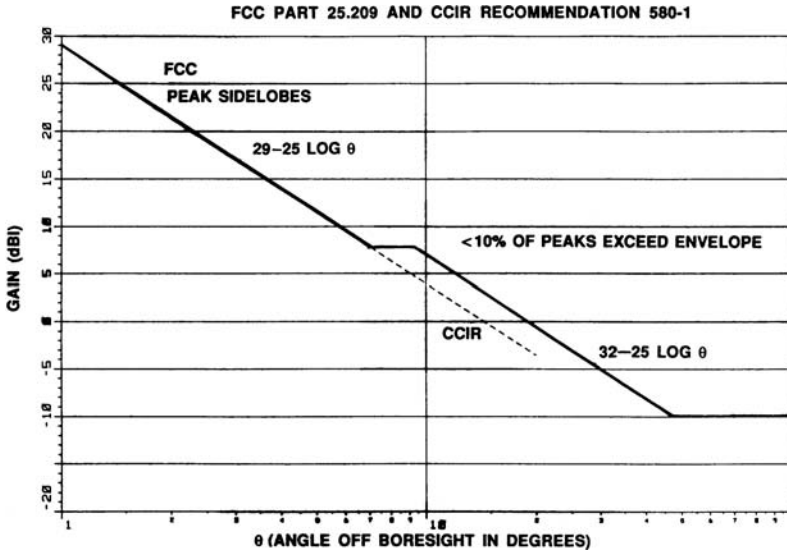


FIGURE 45-1 Sidelobe envelope as defined by CCIR Recommendation 580-1 and the FCC

This envelope takes into consideration the limitations of small-antenna design and is representative of measured patterns of well-designed dual-reflector antennas.

For antennas used for deep-space communications, there is no specific sidelobe requirement. However, since the prime driver for deep-space missions is to receive extremely weak signals over vast interplanetary distances, a key element of telecommunications link performance is the received power signal-to-noise ratio (SNR), which is given by

$$S/N \approx \frac{P_T G_T G_R}{4\pi R^2 N} = \frac{4\pi P_T A_T A_R}{\lambda^2 R^2 k B T_s} \quad (45-1)$$

where

- P_T = spacecraft transmit power
- G_T = gain
- A_T = effective area of the transmit (spacecraft) antenna
- G_R = gain
- A_R = effective areas of the receive ground antennas
- N = total noise
- R = distance to the spacecraft
- T_s = receive system noise temperature
- λ = wavelength
- B = bandwidth
- k = Boltzman's constant

To do its part effectively, the ground antenna system must maximize the ratio of received signal to the receiving system noise power, which is measured by an antenna figure of merit (FM), defined as the ratio of antenna effective area (or equivalent gain) to system noise temperature.

The receive temperature consists primarily of the antenna feed system and amplifier contributions. For assessing the antenna FM, it is desirable to draw an imaginary reference plane between the receiver system and the antenna system, thus placing all noise contributions in one or the other of these categories. If the receiver contribution (including the feed system losses) is given by T_R and the antenna noise contribution by T_A , then the FM will be given by

$$FM = \frac{G_R}{T_R + T_A} \quad (45-2)$$

Antenna noise temperature properties are very significant contributors to FM, especially for cases of low receiver noise temperature systems. Thus, to maximize the FM for a given antenna size and frequency of operation, it is necessary to both maximize the antenna gain and minimize the total system noise temperature. Since the individual contributions to noise temperature are additive and essentially independent of each other, it is necessary to individually minimize each contribution. If any one of the noise contributions is large, minimizing the others only marginally improves the FM. However, by using cryogenic amplifiers, the receiver noise temperature contribution can be small (as low as 2 to 3 K for the best performing masers), and it then becomes imperative to minimize both the antenna and feed system contributions. For ambient conditions, it should be noted that feed system losses contribute to noise temperature at the rate of 7 K per 0.1-dB loss.

Ground station antennas can be grouped into two broad categories: single-beam antennas and multiple-beam antennas. A single-beam antenna is defined as an antenna that generates a single beam that is pointed toward a satellite by means of a positioning system. A multiple-beam Earth station antenna is defined as an antenna that generates multiple beams by employing a common reflector aperture with multiple feeds illuminating that aperture. The axes of the beams are determined by the location of the feeds. The individual beam identified with a feed is pointed toward a satellite by positioning the feed, generally without moving the reflector. An important subclass of single-beam antennas is the beam-waveguide (BWG) antenna, which is composed of one or more feedhorns with a series of flat and curved mirrors arranged so that power can be propagated from the horn through the mirrors to the main reflector with minimum loss.

45.2 SINGLE-BEAM EARTH STATION ANTENNAS

Single-beam antenna types used as Earth stations are paraboloidal reflectors with focal-point feeds (prime-focus antenna), dual-reflector antennas such as the Cassegrain, Gregorian, and dual-shaped configurations, BWG antennas, offset-fed paraboloidal antennas, and offset-fed multiple-reflector antennas. Each of these antenna types has its own unique characteristics, and the advantages and disadvantages have to be considered when choosing one for a particular application.

Axisymmetric Dual-Reflector Antennas

The predominant choice of designers of large Earth station antennas has been the axially symmetrical dual-reflector antenna (Cassegrain or Gregorian, classical or shaped).

1. The classical Cassegrain geometry^{2,3} employs a paraboloidal surface for the main reflector and a hyperboloidal surface for the subreflector (see Figure 45-2). The paraboloidal

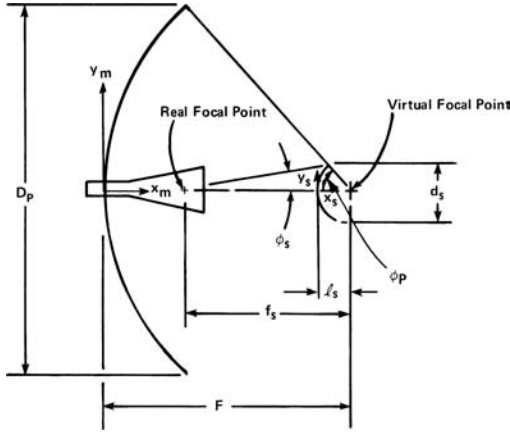


FIGURE 45-2 Geometry of the Cassegrain antenna system

reflector is a point-focus device with a diameter, D_p , and a focal length, f_p . The hyperboloidal subreflector has two foci. For proper operation, one of the two foci is the real focal point of the system and is located coincident with the phase center of the feed; the other focus, the virtual focal point, is located coincident with the focal point of the main reflector. The parameters of the Cassegrain system, as shown in Figure 45-2, are related as follows:

$$\phi_p = 2 \tan^{-1}(0.25D_p / F_p) \quad (45-3)$$

$$f_s / d_s = 0.5(\cot \phi_p + \cot \phi_s) \quad (45-4)$$

$$\ell_s / f_s = 0.5(1 - \{\sin[0.5(\phi_p - \phi_s)] / \sin[0.5(\phi_p + \phi_s)]\}) \quad (45-5)$$

In a typical design, the parameters F_p , D_p , and ϕ_s are chosen, and the remaining three parameters are then calculated. Granet⁴ has given design formulas and procedures for both Cassegrain and Gregorian starting with different design parameters and design considerations such as minimum blockage.

The contours of the main reflector and subreflector are given by

$$\text{Main reflector: } y_m^2 = 4F_p x_m \quad (45-6)$$

$$\text{Subreflector: } (y_s / b)^2 + 1 = (x_s / a + 1)^2 \quad (45-7)$$

where

$$a = (f/2e) \quad b = a\sqrt{e^2 - 1}$$

$$e = \sin[0.5(\phi_p + \phi_s)] / \sin[0.5(\phi_p - \phi_s)]$$

The quantities a , b , and e are half of the transverse axis, half of the conjugate axis, and the eccentricity parameters of the hyperboloidal subreflector, respectively.

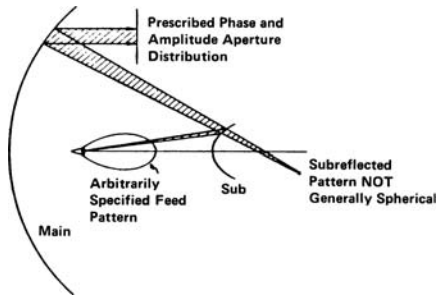


FIGURE 45-3 Circularly symmetric dual-shaped reflectors

2. An extremely important generalization of the Cassegrain geometry is that it consists of a special-shaped quasi-paraboloidal main reflector and a shaped quasi-hyperboloidal subreflector.⁵⁻⁸ Green⁹ observed that in dual-reflector systems with high magnification—essentially a large ratio of main-reflector diameter to subreflector diameter—the distribution of energy (as a function of angle) is largely controlled by the subreflector curvature. The path length or phase front is dominated by the main reflector (see Figure 45-3). Kimber⁵ and Galindo^{6,7} found a method for simultaneously solving for the main-reflector and subreflector shapes to obtain an exact solution for both the phase and the amplitude distributions in the aperture of the main reflector of an axisymmetric dual-reflector antenna. Their technique, based on geometrical optics, involves solving two simultaneous, nonlinear, first-order, ordinary differential equations. Figure 45-4 gives the geometry showing the path of a single ray.

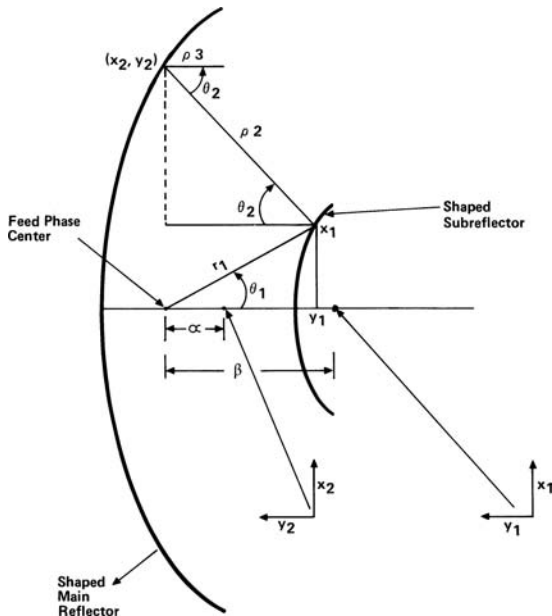


FIGURE 45-4 Dual-shaped reflector geometry

The feed phase center is located as shown, and the feed is assumed to have a power radiation pattern (θ_1). The parameters α and β represent, respectively, the distance of the feed phase center from the aperture plane and the distance between the feed phase center and the back surface of the subreflector. The constraints to the dual-reflector system are as follows:

- a. The phase distribution across the main-reflector aperture plane will be uniform, or

$$r_1 + \rho^2 + \rho^3 + C_p(\theta_1) = \text{constant} \quad (45-8)$$

over the range $0 \leq \theta_1 \leq \theta_{1\max}$. $C_p(\theta_1)$ represents the phase distribution across the primary-feed radiation pattern in units of length.

- b. The feed energy, or ray bundles intercepted and reflected by the subreflector, is conserved and redistributed according to a specified aperture distribution, or

$$I(\theta_1) \sin(\theta_1) d\theta_1 = C \cdot I(X_2) X_2 dX_2 \quad (45-9)$$

where $I(X_2)$ represents the power radiation distribution across the main-reflector aperture, and C represents a constant that is determined by applying the conservation-of-power principle.

$$\int_0^{\theta_{1\max}} I(\theta_1) \sin \theta_1 d\theta_1 = C \int_{X_{2\min}}^{X_{2\max}} I(X_2) X_2 dX_2 \quad (45-10)$$

The lower limit of integration over the main reflector can be arbitrarily chosen so that only an annular region of the main reflector is illuminated.

- c. Snell's law must be satisfied at the two reflecting surfaces; applying it yields

$$\frac{dY_1}{dX_1} = \tan\left(\frac{\theta_1 - \theta_2}{2}\right) \quad (45-11)$$

$$\frac{dY_2}{dX_2} = -\tan\left(\frac{\theta_2}{2}\right) \quad (45-12)$$

Solving Eqs. 45-9 through 45-12 simultaneously results in a nonlinear, first-order differential equation of the form

$$\frac{dY_1}{dX_2} = f(\theta_1, \theta_2, \alpha, \beta, \text{etc.}) \quad (45-13)$$

which leads to the cross sections of each reflector when subject to the boundary condition $Y_1(X_2 = X_{2\max}) = 0$, where X_2 is the independent variable. Equation 45-13 can be solved numerically by using an algorithm such as a Runge-Kutta, order 4.

The above procedure is based on geometrical optics (GO), but it is evident that the assumptions of GO are far from adequate when reflectors are small in terms of wavelengths. An improvement in the design approach is to include the effects of diffraction. Clarricoats and Poulton¹⁰ reported a gain increase of 0.5 dB for a diffraction-optimized design over the GO design with a 400λ -diameter main reflector and 40λ -diameter subreflector.

Prime-Focus-Fed Parabolic Antennas

Prime-focus-fed parabolic reflector antennas are also often employed as Earth station antennas. For moderate to large aperture sizes, this type of antenna has excellent sidelobe performance in all angular regions except the spillover region around the edge of the reflector. The CCIR sidelobe specification can be met with this type of antenna. See Chapter 15 on reflector antennas for more information on parabolic antennas.

Offset-Fed Reflector Antennas

The advantage of offset-fed reflector antennas is the elimination of feed and feed-support blockage. This can be extremely important for radio-telescope antennas. Because they are asymmetric, offset antennas are more complicated to build per unit aperture than circularly symmetric antennas, but the additional cost of the offset aperture generally isn't warranted for telecommunications. Consequently, their use is generally reserved for smaller antennas, where eliminating blockage is important and the cost differential for producing the asymmetric aperture is small.

Offset-fed reflector antennas can employ a single reflector or multiple reflectors, with two-reflector types the more prevalent of the multiple-reflector designs. The offset front-fed reflector, consisting of a section of a paraboloidal surface (see Figure 45-5), minimizes diffraction scattering by eliminating the aperture blockage of the feed and feed-support structure. Sidelobe levels of $(29-25 \log \theta)$ dBi can be expected from this type of antenna (where θ is the far-field angle in degrees) with aperture efficiencies of 65 to 80 percent. The increase in aperture efficiency compared with that of axisymmetric prime-focus-fed antennas is due to the elimination of direct blockage. For a detailed discussion of this antenna, see C. A. Mentzer.¹³

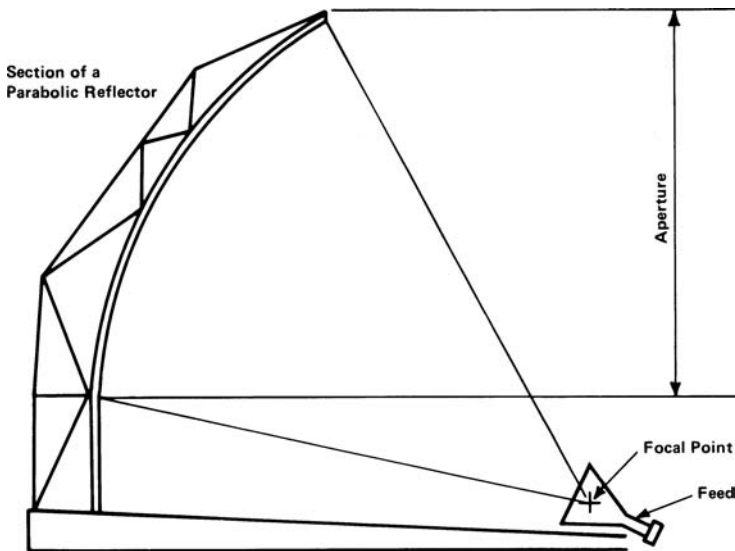


FIGURE 45-5 Basic offset-fed paraboloidal antenna

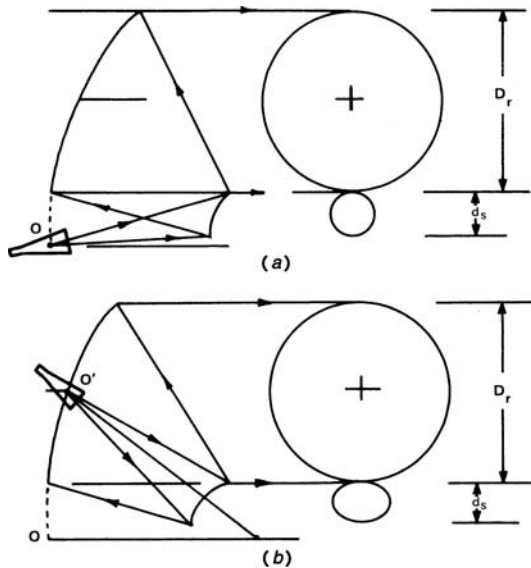


FIGURE 45-6 Offset dual-reflector geometries: (a) double-offset geometry (feed phase center and paraboloidal vertex at 0), and (b) open Cassegrain geometry (feed phase center located at 0; paraboloidal vertex at 0)

Offset-fed dual-reflector antennas exhibit sidelobe performance similar to that of front-fed offset reflectors. Two offset-fed dual-reflector geometries are used for Earth station antennas: the double-offset geometry shown in Figure 45-6a, and the open Cassegrain geometry introduced by Cook et al¹¹ of Bell Laboratories and shown in Figure 45-6b. In the double-offset geometry, the feed is located below the main reflector, and no blocking of the optical path occurs. In contrast, the open Cassegrain geometry is such that the primary feed protrudes through the main reflector; thus, it is not completely blockage-free. Nevertheless, both of these geometries have the capability of excellent sidelobe and efficiency performance.

The disadvantage of offset-fed dual-reflector antennas is that they are asymmetric, which results in increased manufacturing cost. Also, offset-geometry antennas, when used for linear polarization, have a depolarizing effect on the primary-feed radiation and produce two cross-polarized lobes within the main beam in the plane of symmetry. For circular polarization, this depolarization effect introduces a beam squint whose direction is dependent upon the sense of polarization. The beam squint is approximately given by¹² $\psi_s = \arcsin [\lambda \sin(\theta_0)/4\pi F]$, where θ_0 is the offset angle, λ is the free-space wavelength, and F is the focal length. The effects of beam squint versus polarization for circular polarization and the relatively high off-axis cross-polarization performance for linear polarization must be considered by the antenna designer, since these characteristics may present a problem for the overall Earth station operation. Offset dual-reflector geometry has the capability of employing a feed tilt to correct for the polarization problems associated with the offset geometry resulting in the equivalent polarization performance of an axially symmetric antenna.^{13,15}

Galindo-Israel, Mittra, and Cha¹⁶ have studied the offset-shaped dual-reflector geometry for high-aperture-efficiency applications. Their analytical techniques are reported to result in efficiencies in the 80 to 90 percent range.¹⁷ See Galindo-Israel¹⁸ for a discussion of the synthesis of offset-shaped reflector antennas.

Beam-Waveguide Systems

A BWG feed system is composed of one or more feedhorns with a series of flat and curved mirrors arranged so that power can be propagated from the horn through the mirrors to the reflector system with minimum losses. Horns and equipment can then be located in a large, non-tipping enclosure at an accessible location.

Feeding a large, low-noise ground antenna via a beam BWG system has several advantages over placing the feed directly at the focal point of a dual-reflector antenna. For example, significant simplifications are possible in the design of high-power, water-cooled transmitters and low-noise cryogenic amplifiers. These systems do not have to rotate as in a normally fed dual reflector. Also, since a BWG system can transmit power over considerable distances at very low losses, BWG optics is often used in the design of very high-frequency feed systems.

Various design techniques¹⁹ have been used, but most frequently they are designed using either GO or Gaussian beams.

Geometrical Optics

The design for a BWG system using GO is based upon criteria introduced by Mizusawa and Kituregawa^{20,21} that guarantee a perfect image from a reflector pair. Mizusawa's criterion can be briefly stated as follows: For a circularly symmetric input beam, the conditions on a conic reflector pair necessary to produce an identical output beam are

1. The four loci (two of which may be coincident) associated with the two curved reflectors must be arranged on a straight line; and
2. The eccentricity of the second reflector must be equal to the eccentricity or the reciprocal of the eccentricity of the first reflector.

Figures 45-7a to 45-7c show some curved reflector pair orientations that satisfy Mizusawa's criteria.

Mizusawa's criteria were used in the design of NASA's first DSN BWG antenna, which is shown in Figures 45-8 and 45-9 and described in detail in Chapter 7 of *Large Antennas of the Deep Space Network*.¹ The design of the center-fed BWG consists of a beam magnifier ellipse in a pedestal room located below ground level that transforms a 22-dB gain feedhorn into a high-gain 29-dB pattern for input to a Mizusawa four-mirror (two flat and two paraboloid-case) BWG system. The system was initially designed for operation at 8.45 GHz (X-band) and 32 GHz (Ka-band) and has less than a 0.2-dB loss at X-band

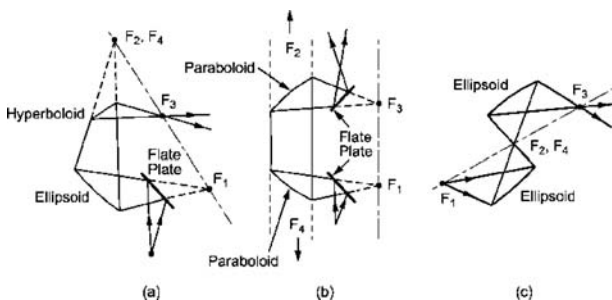


FIGURE 45-7 Examples of two curved-reflector BWG configurations

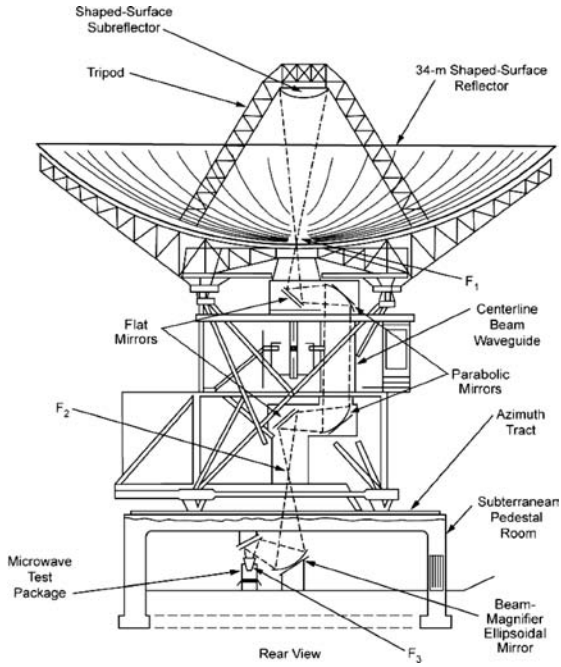


FIGURE 45-8 Typical center-fed beam-waveguide antenna



FIGURE 45-9 First DSN beam-waveguide antenna named DSS-13

(determined by comparing the gain of a 29-dB gain horn feeding the dual-shaped reflector system with that obtained using the BWG system) Gaussian beam.

While GO is useful for designing systems with electrically large mirrors ($>50\lambda$ diameter with -20 dB or greater edge taper), some BWGs may be operated at low frequencies where the mirrors may be as small as 20λ in diameter. Due to diffraction effects, the characteristics of a field propagated between small BWG mirrors ($<30\lambda$ in diameter) will be substantially different from those resulting from the GO solution. For these small-mirror cases, the Gaussian beam technique was used.

The Gaussian beam is an approximate solution to a wave equation describing a beam that is unguided but effectively confined near an axis. The zero-order mode is normally used in the design. A major advantage of the Gaussian technique is the simplicity of the Gaussian formula, which is easy to implement and requires negligible computation time.

Because of the negligible computation time, a Gaussian solution can be incorporated into an optimization routine, providing a convenient method to search design parameters for a specified frequency range, mirror sizes and locations, and horn parameters.

G. Goubau gave the first mathematical expression of Gaussian modes, derived from the solution of Maxwell's equations. These Gaussian modes are described by a continuous spectrum of cylindrical waves.²²

Although Gaussian-beam analysis is fast and simple, it is less accurate than the physical optics (PO) solution for smaller mirrors ($<30\lambda$ diameter). However, by designing with Gaussian-beam analysis, then checking and adjusting using PO analysis, an accurate and efficient tool can be fashioned. Veruttipong²³ developed such a tool for designing a second 34-m BWG antenna for the DSN. The goal was to provide good performance over the range of 2 to 32 GHz.

The design is similar to that of the DSS-13 antenna (see Figure 45-9) in that it uses three curved mirrors (one in the basement room and two rotating in azimuth) and a 34-m dual-shaped reflector antenna (see Figure 45-10). Multiple-frequency operation is provided

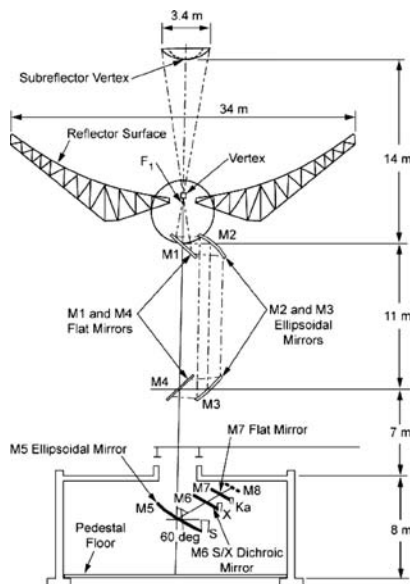


FIGURE 45-10 BWG designed using Gaussian techniques

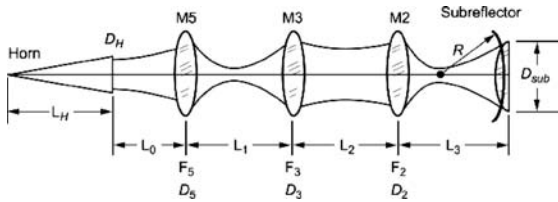


FIGURE 45-11 Parameters for BWG Gaussian design

by the dichroic mirrors. The goal is to have the radius of curvature and -18 -dB beam diameter of the Gaussian beam at the subreflector be the same at all frequencies. The size and locations of the mirrors are relatively fixed because of the basic structure geometry, so the pertinent variables are the horn diameters, horn positions, and mirror curvatures. Approximating the mirrors by applying a thin-lens formula (see Figure 45-11) and by using Gaussian-mode analysis to iterate the various design parameters, a design is achieved that meets the initial design constraint of identical patterns at the subreflector. The 34-m BWG antenna was built and measured at S-band, X-band, and Ka-band (see Chapter 8 of *Large Antennas of the Deep Space Network*¹ for details).

45.3 MULTIPLE-BEAM EARTH STATION ANTENNAS

There has been increasing interest in using a single antenna to receive signals simultaneously from several satellites. This interest has prompted the development of several multibeam-antenna configurations that employ fixed reflectors and multiple feeds. The antenna engineering community has been investigating multibeam antennas for many years. In fact, in the middle of the 17th century, Christian Huygens and Sir Isaac Newton first studied the spherical mirror, and the first use of a spherical reflector as a microwave antenna occurred during World War II. More recently, the spherical-reflector, the torus-reflector, and the dual-reflector geometries, all using multiple feeds, have been offered as antennas with simultaneous multibeam capability. Chu²⁴ in 1969 addressed the multiple-beam spherical-reflector antenna for satellite communication, Hyde²⁵ introduced the multiple-beam torus antenna in 1970, and Ohm²⁶ presented a novel multiple-beam Cassegrain-geometry antenna in 1974. All three of these approaches, as well as variations of scan techniques for the spherical reflector, are discussed below.

Spherical Reflectors

The properties, practical applications, and aberrations of spherical reflectors are not new to microwave-antenna designers. The popularity of these reflectors is primarily due to the large angle through which the radiated beam can be scanned by translation and orientation of the primary feed. This wide-angle property results from the symmetry of the surface. Multiple-beam operation is realized by placing multiple feeds along the focal surface. In the conventional use of the reflector surface, the minimum angular separation between adjacent beams is determined by the feed-aperture size. The maximum number of beams is determined by the percentage of the total sphere covered by the reflector. In the alternative configuration described below, these are basically determined by the f/D ratio and by the allowable degradation in the radiation pattern.

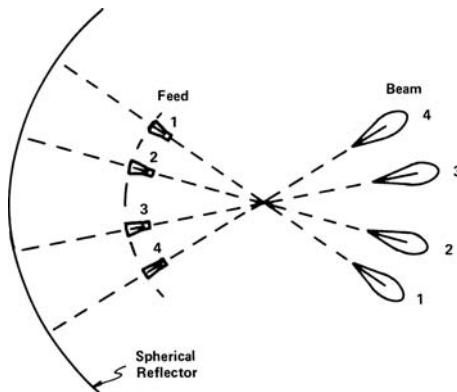


FIGURE 45-12 Conventional spherical multibeam antenna using extended reflector and multiple feeds

In the conventional use of spherical reflectors, the individual feed illuminates a portion of the reflector surface so that a beam is formed coincident to the axis of the feed. The conventional multibeam geometry is shown in Figure 45-12. All the beams have similar radiation patterns and gains, although the performance of spherical reflectors is degraded in comparison with the performance of a paraboloid. The advantage of this design is that the reflector area illuminated by the individual feeds overlaps, reducing the surface area for a given number of beams in comparison with individual single-beam antennas.

The alternative multibeam-spherical-reflector geometry is shown in Figure 45-13. For this geometry, each of the feed elements points toward the center of the reflector, with the beam steering accomplished by the feed position. This method of beam generation leads to considerable increase in aberration, including coma; therefore, the radiation patterns of the off-axis beams are degraded with respect to the on-axis beam. This approach does not take advantage of the spherical-reflector properties that exist in the conventional approach. In fact, somewhat similar results could be achieved using a parabolic reflector with a large f/D .

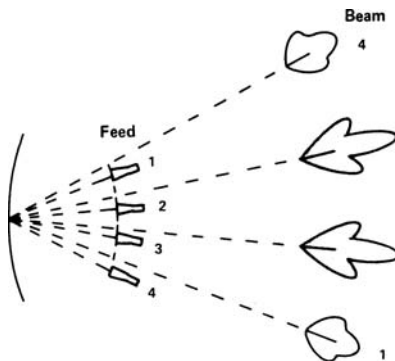


FIGURE 45-13 Alternative spherical multibeam antenna using minimum reflector aperture with scanned beam feeds

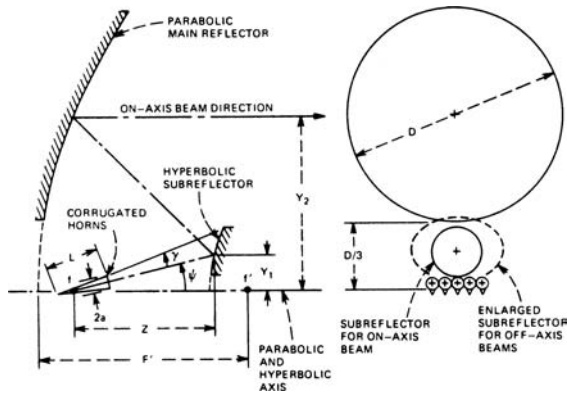


FIGURE 45-15 Geometry of the offset-fed multibeam Cassegrain antenna (© American Telephone and Telegraph Company 1974. Reprinted by permission.)

is compact and has a large focal-length-to-diameter ratio (f/D), which reduces aberrations to an acceptable level even when the beam is moderately far off axis. Low-sidelobe performance is achieved by using a corrugated feedhorn that produces a Gaussian beam.

A typical antenna design consisting of a 10-m projected aperture would yield half-power beamwidths (HPBW) and gain commensurate with an axisymmetric 10-m antenna, 0.5 HPBW, and 51-dB gain at 4 GHz. The subreflector would need to be approximately a 3-m by 4.5-m elliptical aperture. The feed apertures would be approximately 0.5 m in diameter. The minimum beam separation would be less than 2 deg, which is more than sufficient to allow use with synchronous satellites with orbit spacing of 2 deg or greater. For the ± 15 -deg scan, the gain degradation would be approximately 1 dB, and the first sidelobe would be approximately 20 to 25 dB below the main-beam peak.

45.4 ANGLE-TRACKING TECHNIQUES

Automatic tracking of satellite position is required for many Earth station antennas. Monopulse, conical-scan, and sequential-lobing techniques described in the chapter on tracking antennas are applicable for this purpose. Two other types of tracking techniques are also commonly employed: single-channel monopulse (pseudo monopulse) and step-track. Single-channel monopulse is a continuous angle-tracking scheme, and steptrack is a time-sequencing, signal-peaking technique.

Single-Channel Monopulse Technique

This technique uses multiple elements or modes to generate a reference signal, an elevation error signal, and an azimuth error signal. The two error signals are then combined in a time-shared manner by a switching network that selects one of two phase conditions (0 and 180 deg) for the error signal. The error signal is then combined with the reference signal, with the resultant signal then containing angle information. This allows the use of a single receiver for the tracking channel. This technique is equivalent to sequential lobing in which the lobing is done electrically and can be adjusted to any desired fixed or variable scan rate.

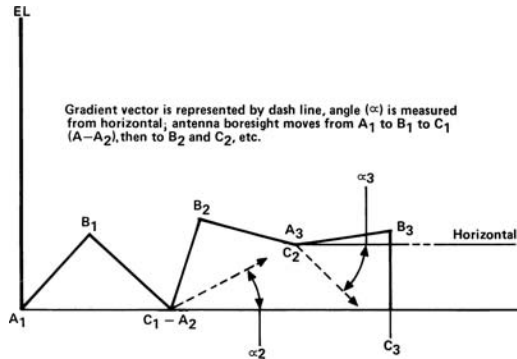


FIGURE 45-16 Signal-strength gradients

Steptrack Technique

Steptracking is a technique employed primarily for maintaining the pointing of an Earth station's antenna beam toward a geosynchronous satellite. Since most geosynchronous orbiting satellites only slowly traverse a small angular region, a simple peaking technique can be used to keep the Earth station beam correctly pointed. The signal-peaking, or steptrack, routine is a software technique that maneuvers the antenna toward the signal peak by following a path along the steepest signal-strength gradient or by using some other algorithm that accomplishes the same result.^{28,29} Any method of direct search for a maximization of signal is applicable; one such method calculates the local gradient by determining the signal strength at three points (A, B, C ; see Figure 45-16). From the signal strength at points A, B , and C , an angle α_2 , representing the unit's gradient vector angle relative to the azimuth axis, is computed. Once C_1 has been determined, A_2 is set equal to C_1 and another angle θ_2 is defined by $\theta_2 = \theta_1 + \alpha_2$ ($\theta_1 = 0$). The angle is used to relate pedestal movements to a fixed-pedestal coordinate system. The procedure is repeated a number of times until the antenna boresight crosses throughout the peak. At this time, the step size is reduced by one-half, and for each time that the peak is crossed it is again reduced by one-half until peaking resolution is accomplished.

45.5 POLARIZATION

Many satellite communications systems are dual-polarized (frequency reuse) and are therefore susceptible to interference from the intended cross-polarized signals when the medium of propagation is such that the ellipticity of the signals is changed. This condition can occur during atmospheric conditions such as rain and Faraday rotation in the ionosphere. Therefore, in order to maintain sufficient polarization discrimination, special precautions must be taken in regard to the polarization purity of dual-polarization antennas; indeed, adaptive polarization-correcting circuits may be necessary.^{30,31}

Several types of polarization-discrimination-enhancement schemes may be used. They include:

- Simple rotation of the polarization major-ellipse axis to correct for rotation due to the ionosphere (applicable for linear polarization). Transmit and receive rotations are in opposite directions.

- Suboptimal correction of depolarizations; ellipticity correction with respect to one of the signals but not orthogonalizing the two signals (differential phase).
- Complete adaptive correction including orthogonalization (differential phase and amplitude).

Polarization-enhancement schemes may be implemented at RF frequencies, or they may follow a down-conversion stage and be implemented at intermediate frequencies (IFs). In either case, the circuitry must operate over the full bandwidth of the communications channel. Kreutel et al³¹ treat the implementation of RF frequencies in detail. See Gianatasio^{32,33} for both RF and IF circuitry and experiments to verify performance of the circuitry.

45.6 MEASUREMENT TECHNIQUES

A number of sophisticated measurement techniques are required to properly characterize the performance of a ground station antenna. There is a need to accurately measure the main reflector surface, which is achieved either by microwave holography, theodolite measurements, or photogrammetry. To determine an antenna's FM, it is necessary to measure gain (or efficiency) and noise temperature.

Microwave Holography

Microwave holography^{34–36} has proven to be a valuable technique for characterizing the mechanical and RF performance of large reflector antennas. This technique uses the Fourier transform relation between the far-field radiation pattern of an antenna and the aperture field distribution. Resulting aperture phase and amplitude data can be used to characterize various performance parameters, including panel alignment, panel shaping, subreflector position, antenna aperture illumination, and gravity deformation effects.

The raw data (the observable) for this technique is the complex far-field pattern of the antenna under test. For large reflector antennas, geostationary satellites are commonly used as far-field signal sources. A typical measurement system consists of the antenna under test, a reference antenna (generally much smaller), and a narrow-bandwidth receiver system architecture used for receiving the narrow-band beacon signal from the satellite. The reference antenna is needed as a phase reference for the measurement as well as to keep the receiver in phase lock with the carrier.^{37,38} The mathematical details are given in Rochblatt.³⁵ Essentially, the inverse Fourier transform of the far-field radiation pattern yields a phase map of the aperture distribution which, through GO, can be related to the surface error.

Accuracy across the holographic map varies with aperture amplitude illumination. Accuracy is better at the center of the dish and gradually worsens toward the edge, where the illumination falls off rapidly. For a uniformly illuminated antenna design, accuracy remains relatively constant, becoming quickly worse just at the edge, also where the illumination falls off rapidly. Typically, accuracies of 0.05 mm to 0.10 mm were achieved with a corresponding SNR of 73 dB to 60 dB.

The resulting aperture phase function needs to be corrected for modulo- 2π phase errors. These occur due to the small measurement wavelength and the large phase error that is partially caused by pointing and subreflector position errors. Next, it is necessary to remove those phase errors caused by pointing and subreflector position errors; this is achieved using the global best-fit paraboloid. Here, the entire data set is weighted-least-squares fitted to the paraboloid, permitting six degrees of freedom in the model, three vertex translations,

two rotations, and a focal length change. The antenna surface axial rms error is then computed with respect to the position of the fitting paraboloid.

The resultant aperture function at the end of this process is defined here as the effective map since it includes all phase effects that are contributing to antenna performance. These frequency-dependent effects consist of the subreflector scattered (frequency-dependent) feed phase function and struts' diffraction effects. Removal of the feed phase function and subreflector-support structure diffraction effects results in a frequency-independent map, which is defined here as the mechanical map. In general, panel-setting information is derived from the mechanical map since the latter is frequency-independent, and the antenna needs to be operated over a range of frequencies. Panel-setting information is determined by sorting all the data points within each panel and performing a rigid-body least-squares fit. The panel-setting algorithm allows for one translation and two rotations.

Theodolite Measurements

Theodolite measurements are typically used in the initial alignment of the antenna, after installation of the reflector panels. The method of determining the reflector surface is based on angular measurements performed with a theodolite.^{39,40} The theodolite is located at the main reflector vertex and is used to measure the angles to targets installed at the individual reflector panel corners. After installation on the antenna, the panels are drilled to accept targets by using a metal strap gage to position the holes.

Errors in theodolite measurements are caused by uncertainty in the angular measurement of the target position and radial distance to the target. At a radius of 13 m, which is the surface area median radius for a 34-m antenna, a 0.001-deg error in the angle to the target corresponds to a 0.2-mm error normal to the surface. Similarly, a 1.0-mm error in the radial distance corresponds to a 0.26-mm error normal to the surface.

A typical error budget⁴⁰ for the surface of a 34-m antenna is 0.3 mm root mean square (RMS). By far, the greatest source of error is in the radial distance measurement provided by using a drill tape. If, instead, a laser range finder is used to measure radial distance, the error could be reduced from 0.3 mm rms to 0.2 mm rms. In any case, microwave holography has been used to characterize and reset the main reflector surface after initial operation using theodolite-set panels.

A Leica TDM-5000 total station theodolite can also be used to measure large reflectors.⁴¹ It calculates vertical and horizontal angles and distances, downloads them to a PC that converts the coordinates from the spherical to a Cartesian system, and can be used to command the instrument to motor to a desired look angle.

Photogrammetry

Photogrammetry, as its name implies, is a three-dimensional coordinate-measuring technique that uses photographs as the fundamental medium for metrology (measurement). The fundamental principle used by photogrammetry is triangulation. By taking photographs from at least two different locations, so-called "lines of sight" can be developed from each camera to points on the object. These lines of sight (sometimes called rays, owing to their optical nature) are mathematically intersected to produce the three-dimensional coordinates of the points of interest. Photogrammetry can be used to measure the surface of a reflector antenna.⁴² Reflective targets are placed uniformly over the surface of the dish and photographs are taken from several different angles. With the aid of calibration targets, the target data are processed to produce a map of the reflective surface.

Aperture Gain and Efficiency Measurements

Determination of aperture efficiency follows a methodology developed by the Jet Propulsion Laboratory (JPL).^{43,44} For a radio source of known flux density, the increase in system noise temperature as determined by the boresight measurements is a measurement of antenna efficiency. Peak radio source noise temperatures are measured using a boresight algorithm.⁴⁵ These measured quantities are then normalized by dividing by T_{100}/C_r , where T_{100} is 100 percent antenna efficiency temperature, and C_r is the source size correction factor.^{43,46} Values of the temperature $T_a = T_{100}/C_r$ for radio sources used to calibrate DSN antennas are given in Richter.⁴⁷ The effects of atmospheric attenuation must also be considered, so the final measured values are then corrected to yield vacuum conditions. The boresight technique, determining antenna efficiency, and the correction for atmospheric attenuation are described in more detail in the following sections.

Boresight Technique The typical method used for calibrating the pointing of DSN antennas is the seven-point boresight technique, which moves the antenna sequentially in both the cross-elevation (XEL) and elevation (EL) directions, both on and off the source. In each direction, the antenna is positioned off-source 10 half-power (one-sided) beamwidths; 1 half-power beamwidth at the 3-dB point; approximately 0.576 half-power beamwidth at the 1-dB point; on source; and then similar offsets on the other side. For example, at X-band with the 34-m antenna, the full 3-dB beamwidth is 65 mdeg. The offsets used are 325, 32.5, 18.7, and 0 mdeg in each direction. At Ka-band, with a full 3-dB beamwidth of 17 mdeg, the offsets were 85, 8.5, 4.9, and 0 mdeg.

For each scan, an off-source baseline is generated from the two off-source points. A Gaussian curve (relative to the baseline) is fitted to the five remaining on-source points, and the peak value of the curve is calculated. Also, the position of the peak is calculated as a measure of the pointing error for that scan. One pair of scans (one XEL and one EL) is considered to be one measurement or data point. The seven pointing offsets for each new scan are corrected for pointing errors found from the previous similar scan in order to maintain pointing throughout a track.

Antenna Efficiency As defined here, the efficiency is referenced to the input of the low-noise amplifier and includes the losses of the feed system. Alternative methods specify efficiency at the aperture of the feedhorn or at the antenna aperture itself. The term “aperture efficiency” refers to antenna gain relative to that of a uniformly illuminated circular aperture having the same diameter as the antenna, e.g., a 70 percent efficient antenna has 1.549-dB less peak gain than does a uniformly illuminated circular aperture.

Radio-source noise temperature increase measured by the antenna is given by⁴⁸:

$$\Delta T = \frac{\eta SA}{2kC_r C_p} \quad (45-14)$$

where

η = antenna aperture efficiency

S = radio source flux, W/m²/Hz

A = antenna area, m²

k = Boltzmann constant, $1.3806503 \times 10^{-23}$ J/K

C_r = source size correction, typically 1.0 for point sources, up to ~1.5 for extended sources, including planets

C_p = pointing correction, assumed = 1.0

The lux, S , is typically given in units of Janskys, where 1 Jansky (Jy) = 1×10^{-26} W/m²/Hz

Measured ΔT is compared with the quantity $[(SA)/(2kC_r)]$ to give antenna efficiency

Thus:

$$\begin{aligned}\eta &= \frac{\Delta T}{SA / 2kC_r} \\ &= \frac{\Delta T}{T100 / C_r}\end{aligned}\quad (45-15)$$

$T100$ is what would be measured by a perfect antenna looking at a point source emitting the same flux as the observed radio source. C_r is a function of the source structure at a particular frequency and the antenna pattern of a particular antenna at that frequency. For a particular antenna, frequency, and radio source, $T100/C_r$ can thus be specified.

Correction for Atmospheric Attenuation ΔT is an on-off source measurement, and Earth's atmosphere attenuates the true source contribution that would be measured under vacuum conditions. The total atmospheric attenuation is estimated from surface weather conditions during all measurements. The surface temperature, pressure, and relative humidity at the site are recorded every half hour. Typical zenith values of attenuation at Goldstone under average clear-sky conditions are

$$\text{X-band: } A_{\text{zen}} = 0.035 \text{ dB}$$

$$\text{Ka-Band: } A_{\text{zen}} = 0.115 \text{ dB}$$

The attenuation in decibels at elevation angle θ is modeled as

$$A(\theta) = \frac{A_{\text{zen}}}{\sin(\theta)}$$

The loss factor at that elevation angle is

$$L(\theta) = 10^{A(\theta)/10}$$

The “vacuum ΔT ” then becomes

$$\Delta T = L(\theta) \Delta T_{\text{measured}}$$

Noise Temperature Measurements

There are a variety of noise temperature measurement techniques and instruments used in the DSN, such as a total power radiometer, noise-adding radiometer, thermal noise standards (cryogenically cooled, ambient, and hot), all well described in Stelzried.⁴⁹ These instruments typically calibrate low-noise systems by injecting a known amount of noise from either a gas tube or solid-state noise diode into the input. The system output power ratio is then used to calculate the operating noise temperature. However, it can be difficult to determine the equivalent noise temperature of the noise source, as defined at the receive input, because of component and transmission line loss between the noise source and the receiver. Determining the amount of noise injected involves either evaluating the noise source and loss separately or measuring the excess noise temperature directly at the receiver input.

A measurement technique that eliminates the difficulty of calibrating noise temperature and is most often used for low-noise ($\ll 300$ K) receiving systems in the DSN is the ambient termination technique,⁴⁹ in which the operating noise-temperature calibrations are performed by alternately connecting the low-noise amplifier input with a waveguide switch between the antenna and an ambient termination. The same effect can be achieved by alternately placing and removing an absorber load over the front of the feedhorn.

The operating noise temperature of a receiving system can be expressed as

$$T_{\text{op}} = T_a + T_e \quad (45-16)$$

where

T_a = antenna temperature, K

T_e = receiver effective noise temperature, K

Figure 45-17 shows a simplified block diagram of the instrumentation for operating noise-temperature calibrations. The precision attenuator is adjusted for equal output power when the receiver input is alternately connected to the ambient load and the antenna. The power ratio Y for an individual measurement is

$$Y_{\text{ap}}(i) = \frac{T_p + T_e}{T_{\text{op}}(i)} \quad (45-17)$$

where

T_p = ambient termination physical temperature, K

T_e = receiver effective noise temperature, K

$T_{\text{op}}(i)$ = individual operating noise-temperature measurement (on antenna), K

Rearranging yields an expression for $T_{\text{op}}(i)$ in terms of the power ratio Y :

$$T_{\text{op}}(i) = \frac{T_p + T_e}{Y_{\text{ap}}(i)} \quad (45-18)$$

A series of measurements is made, and a best estimate in a least-squares sense of Y_{ap} is obtained.

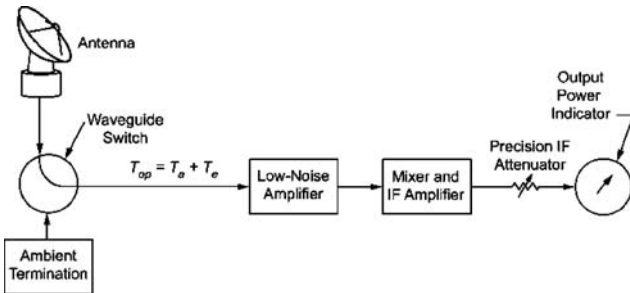


FIGURE 45-17 Simplified block diagram of the operating-noise-temperature calibration instrumentation

The operating noise temperature is then given by

$$T_{\text{op}} = \frac{T_p + T_e}{Y_{\text{ap}}} \quad (45-19)$$

The ambient termination temperature is measured, and the receiver temperature is estimated or calibrated. It is not necessary to know T_e accurately if $T_e \ll T_p$.

In the past, the waveguide beyond cutoff attenuator was the most accurate method to measure Y factors, so the equations for calculating noise temperature were written in terms of Y factors. Most recently, the power meter has been more accurate, and more convenient, mating easily with computers. The method described in Stelzried and Klein^{50,51} does not use Y factors but deals directly with power meter readings, in watts. In addition, this formulation leads to the determination of radiometer linearity.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

1. W. A. Imbriale, *Large Antennas of the Deep Space Network* (Hoboken, NJ: John Wiley & Sons, 2003).
2. P. W. Hannan, "Microwave Antennas Derived from the Cassegrain Telescope," *IRE Trans. Antennas Propagat.*, vol. AP-9 (March 1961): 140–153.
3. P. A. Jensen, "Designing Cassegrain Antennas," *Microwave J.* (December 1962): 10–16.
4. C. Granet, "Designing Axially Symmetric Cassegrain or Gregorian Dual-Reflector Antennas from Combinations of Prescribed Geometric Parameters," *IEEE Antennas Propag.*, vol. 40 (April 1998): 76–82.
5. B. Y. Kinber, "On Two Reflector Antennas," *Radio Eng. Electron. Phys.*, vol. 6 (June 1962): 914–921.
6. V. Galindo, "Design of Dual Reflector Antenna with Arbitrary Phase and Amplitude Distributions," PTGAP Int. Symp., Boulder, July 1963.
7. V. Galindo, "Synthesis of Dual Reflector Antennas," Elec. Res. Lab. Rep. 64-22, University of California, Berkeley, July 30, 1964.
8. W. F. Williams, "High Efficiency Antenna Reflector," *Microwave J.* (July 1965): 79–82.
9. K. A. Green, "Modified Cassegrain Antenna for Arbitrary Aperture Illumination," *IEEE Trans. on Antennas and Propagat.*, vol. AP-11, no. 5 (September 1963): 589–590.
10. P. J. B. Clarricoats and G. T. Poulton, "High-Efficiency Microwave Reflector Antennas—A Review," *IEEE Proc.*, vol. 65, no. 10 (October 1977): 1470–1504.
11. J. S. Cook, E. M. Elam, and H. Zucker, "The Open Cassegrain Antenna: Part I. Electromagnetic Design and Analysis," *Bell Syst. Tech. J.* (September 1965): 1255–1300.
12. A. W. Rudge and N. A. Adatia, "Offset-Parabolic-Reflector Antennas: A Review," *P. IEEE*, vol. 66, no. 12 (December 1978): 1592–1611.
13. H. Tanaka and M. Mizusawa, "Elimination of Cross-Polarization in Offset Dual Reflector Antennas," *Elec. Commun. (Japan)*, vol. 58 (1975): 71–78.
14. Y. Mizuguch, M. Akagawa, and H. Yokoi, "Offset Dual Reflector Antenna," *IEEE T. Int. Symp. Antennas Propag.* (October 1976).

15. C. Dragone, "Offset Multireflector Antennas with Perfect Pattern Symmetry and Polarization Discrimination," *Bell Syst. Tech. J.*, vol. 57 (September 1978): 2663–2684.
16. V. Galindo-Israel, R. Mittra, and A. Cha, "Aperture Amplitude and Phase Control of Offset Dual Reflectors," USNC/URSI & IEEE Antennas Propagat. Int. Symp., May 1978.
17. A. Cha and D. A. Bathker, "Preliminary Announcement of an 85-Percent Efficient Reflector Antenna," *IEEE T. Antennas Propag.*, vol. AP-31, no. 2 (March 1983): 341–342.
18. V. Galindo-Israel, W. A. Imbriale, and R. Mittra, "On the Theory of the Synthesis of Single and Dual Offset Reflector Antennas," *IEEE T. Antennas Propag.*, vol. AP-35, no. 8 (August 1987): 887–896.
19. W. A. Imbriale, "Design and Applications of Beam Waveguide Systems," *1997 IEEE Aerospace Conference Proc.*, vol. 3 (February 1–8, 1997): 121–134.
20. M. Mizusawa and T. Kitsuregawa, "A Beam-waveguide Feed Having a Symmetric Beam for Cassegrain Antennas," *IEEE T. Antennas Propag.*, vol. AP-21 (November 1973): 844–846.
21. T. Veruttipong, J. R. Withington, V. Galindo-Israel, W. A. Imbriale, and D. Bathker, "Design Considerations for Beam-waveguide in the NASA Deep Space Network," *IEEE T. Antennas Propag.*, vol. AP-36 (December 1988): 1779–1787.
22. G. Goubau and F. Schwering, "On the Guide Propagation of Electromagnetic Wave Beams," *IRE T. Antennas Propag.*, vol. AP-9, no. 5 (May 1961): 248–256.
23. W. Veruttipong, J. C. Chen, and D. A. Bathker, "Gaussian Beam and Physical Optics Iteration Technique for Wideband Beam Waveguide Feed Design," *TDA Progress Rep. 42-105*, Jet Propulsion Laboratory, California Institute of Technology (May 15, 1991): 128–135.
24. T. S. Chu, "A Multibeam Spherical Reflector Antenna" *IEEE Antennas Propag. Int. Symp.: Program & Dig.* (December 9, 1969): 94–101.
25. G. Hyde, "A Novel Multiple-Beam Earth Terminal Antenna for Satellite Communication," *Int. Conf. Comm. Rec.*, Conf. Proc. Pap. 70-CP-386-COM (June 1970): 38-24–38-33.
26. E. A. Ohm, "A Proposed Multiple-Beam Microwave Antenna for Earth Stations and Satellites," *Bell Syst. Tech. J.* vol. 53 (October 1974): 1657–1665.
27. G. Hyde, R. W. Kreuter, and L. V. Smith, "The Unattended Earth Terminal Multiple-Beam Torus Antenna," *COMSAT Tech. Rev.*, vol. 4, no. 2 (1974): 231–264.
28. M. J. D. Powell, "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," *Comput. J.*, vol. 7 (1964): 155.
29. J. Kowalik and M. R. Osborne, *Methods for Unconstrained Optimization Problems* (New York: American Elsevier Publishing Company, Inc., 1968).
30. T. S. Chu, "Restoring the Orthogonality of Two Polarizations in Radio Communication Systems, I and II," *Bell Syst. Tech. J.*, vol. 50, no. 9 (November 1971): 3063–3069; vol. 52, no. 3 (March 1973): 319–327.
31. R. W. Kreutel, D. F. DiFonzo, W. J. English, and R. W. Gruner, "Antenna Technology for Frequency Reuse Satellite Communications," *IEEE Proc.*, vol. 65, no. 3 (March 1977): 370–378.
32. A. J. Gianatasio, "Broadband Adaptively-Controlled Polarization-Separation Network," *Seventh Europ. Microwave Conf. Proc.* (September 1977): 317–321.
33. A. J. Gianatasio, "Adaptive Polarization Separation Experiments," final report, NASA CR-145076, (November 1976).
34. D. J. Rochblatt and B. L. Seidel, "DSN Microwave Antenna Holography," *TDA Progress Rep. 42-76*, vol. October–December 1983 (February 15, 1984): 27–42.
35. D. J. Rochblatt, "A Microwave Holography Methodology for Diagnostics and Performance Improvement for Large Reflector Antennas," *TDA Progress Rep. 42-108*, vol. October–December 1991 (February 15, 1992): 235–252.
36. D. J. Rochblatt and B. L. Seidel, "Performance Improvement of DSS-13 34-meter Beam-Waveguide Antenna Using the JPL Microwave Methodology," *TDA Progress Rep. 42-108*, vol. October–December 1991 (February 15, 1992): 253–270.
37. D. J. Rochblatt, "System Analysis for DSN Microwave Antenna Holography," *TDA Progress Rep. 42-97*, vol. January–March 1989 (May 15, 1989): 132–157.

38. D. J. Rochblatt and Y. Rahmat-Samii, "Effects of Measurement Errors on Microwave Antenna Holography," *IEEE T. Anten. Propag.*, vol. 39, no. 7 (July 1991): 933–942.
39. R. Levy, *Structural Engineering of Microwave Antennas* (New York: IEEE Press, 1996): 289–296.
40. R. Levy, "DSS-13 Antenna Structure Measurements and Evaluation," JPL D-8947 (internal document), Jet Propulsion Laboratory, Pasadena, CA (October 1, 1991).
41. M. Brenner, M. J. Britcliffe, W. A. Imbriale, "Gravity Deformation Measurements of 70m Reflector Surfaces," AMTA 2001 Conference, Denver, Colorado, October 21–26, 2001.
42. M. Doneus, "Introduction to Photogrammetry," <http://www.univie.ac.at/Luftbildarchiv/wgv/intro.htm>, accessed September 1, 2006.
43. M. J. Britcliffe, L. S. Alvarez, D. A. Bathker, P. W. Cramer, T. Y. Otoshi, D. J. Rochblatt, B. L. Seidel, S. D. Slobin, S. R. Stewart, W. Veruttipong, and G. E. Wood, "DSS 13 Beam Waveguide Antenna Project: Phase 1 Final Report," JPL D-8451 (internal document), Jet Propulsion Laboratory, Pasadena, CA (May 15, 1991).
44. S. D. Slobin, T. Y. Otoshi, M. J. Britcliffe, L. S. Alvarez, S. R. Stewart, and M. M. Franco, "Efficiency Calibration of the DSS 13 34-meter Beam-Waveguide Antenna at 8.45 and 32 GHz," *TDA Progress Rep. 42-106*, vol. April–June 1991 (August 15, 1991): 283–297.
45. L. S. Alvarez, "Analysis and Applications of a General Boresight Algorithm for the DSS 13 Beam-Waveguide Antenna," *TDA Progress Rep. 42-111*, vol. July–September 1992 (November 15, 1992): 48–61.
46. P. H. Richter and S. D. Slobin, "DSN 70-meter Antenna X- and S-band Calibration Part I: Gain Measurements," *TDA Progress Rep. 42-97*, vol. January–March 1989 (May 15, 1989): 314–351.
47. P. Richter, "DSN Radio Source List for Antenna Calibration," JPL D-3801 (internal document), Rev. C, Jet Propulsion Laboratory, Pasadena, CA (August 19, 1993).
48. C. Stelzried, *The Deep Space Network—Noise Temperature Concepts, Measurements, and Performance*, JPL Publication 82-33, Jet Propulsion Laboratory, Pasadena, CA (September 15, 1982).
49. C. T. Stelzried, "Operating Noise-Temperature Calibrations of Low-Noise Receiving Systems," *Microwave J.*, vol. 14, no. 6 (June 1971): 41.
50. C. T. Stelzried and M. J. Klein, "Precision DSN Radiometer Systems: Impact on Microwave Calibrations," *P. IEEE* (May 1994): 636.
51. C. T. Stelzried and M. Klein, "Corrections to 'Precision DSN Radiometer Systems: Impact on Microwave Calibrations,'" *P. IEEE* (August 1996): 1187.

Chapter 46

Seeker Antennas

James M. Schuchardt

Consultant, Jasper, GA

Dennis J. Kozakoff

USDigiComm, Alpharetta, GA

CONTENTS

46.1	INTRODUCTION	46-2
46.2	SEEKER-ANTENNA ELECTRICAL CONSIDERATIONS	46-2
46.3	IMPACT OF AIRFRAME ON ANTENNA DESIGN	46-4
46.4	SEEKER-ANTENNA MECHANICAL CONSIDERATION	46-5
46.5	APERTURE TECHNIQUES FOR SEEKER ANTENNAS	46-5
46.6	PACKAGING SEEKER ANTENNAS	46-9
46.7	SMALL ARRAYS FOR SEEKER ANTENNAS	46-10
46.8	RADOME EFFECTS	46-11
46.9	EVALUATION OF SEEKER ANTENNAS	46-13
46.10	ADDITIONAL DEVELOPMENTS AND SOURCES OF INFORMATION	46-13

46.1 INTRODUCTION

This chapter will focus on antennas forward-mounted in a missile functioning in the role of seeker of target emissions by radar homing or antiradiation homing. Such a seeker antenna is a critical part of the entire airborne guidance system, which includes the missile radome, seeker antenna, radio frequency (RF) receiver, antenna gimbal, autopilot, and airframe. The seeker has several functions: to receive and track target emissions so as to measure line of sight and/or line-of-sight angular rate, to measure closing velocity, and to provide steering commands to the missile autopilot and subsequently to the control surfaces. The signals received by the missile-mounted seeker antenna or antennas are thus utilized in a closed-loop servocontrol system to guide the missile to the target.

The seeker RF elements, radome and antenna, initially discriminate in angle through the seeker antenna's pencil beam or equivalent. This beam can be steered mechanically (the whole antenna moves), electromechanically (an antenna element such as a subreflector moves), or electronically (there is a phased-array movement). In some situations, the antenna is stationary and only forward-looking; thus beam motion occurs only if the entire missile airframe rotates. Fixed seeker-antenna beams are used when a pursuit navigation (or a variant) guidance algorithm is used. The use of movable seeker antenna beams occurs when a proportional navigation (or a variant) guidance algorithm is used.

46.2 SEEKER-ANTENNA ELECTRICAL CONSIDERATIONS

Seekers generally fall into one of three categories: conical scan (conscan), monopulse, or sequential lobing. Implementation details for each of these are well known.¹ Conscan radars locate the center of the target through the nutation of a single feed (or antenna) that is displaced from the axis. Monopulse antennas most often generate four simultaneous beams that are summed and differenced to produce azimuth and elevation-plane error voltages that are proportional to target displacement from the axis. Sequential-lobing radars sequentially sample the energy in four beams each corresponding to an antenna feed; the error voltage is derived by comparing the amplitude of the signal in opposite feeds.

Conical scan is used primarily for very-short-range, simple guidance systems. Conscan is in limited use because of its vulnerability to a form of ECM known as spin-frequency jamming, in which the jammer reradiates the seeker transmission, amplitude modulated at the conical scan frequency, thus creating spurious angle-error signals.² Sequential-lobing radars are seldom used in today's high-performance missile guidance systems, giving preference to monopulse, which provides simultaneous lobing of both principal-plane error channels.

Modern polarimetric seeker radars require dual-polarized seeker-antenna performance.³ Other parameters of particular concern include instantaneous bandwidth, transmit power capability, and low antenna sidelobe performance.⁴

Seeker antennas can be realized in a variety of technologies ranging from microstrip to waveguide (see Figure 46-1). Table 46-1 presents a summary of commonly employed seeker antenna

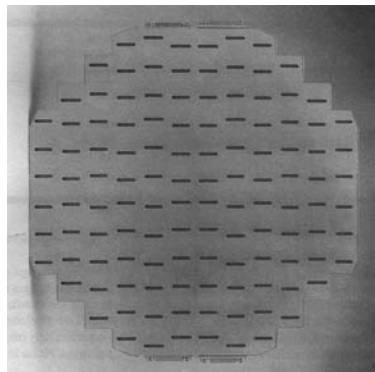


FIGURE 46-1 Representative planar-seeker waveguide flat-plate antenna (*Courtesy of Rantec Microwave & Electronics*)

TABLE 46-1 Seeker-Antenna Radiator Summary

Basic Radiator	Bandwidth Potential	Polarization	Dual-polarization Capability	Power
Stripline	Narrow	Linear or circular	Yes	Intermediate
Microstrip	Narrow	Linear or circular	Yes	Intermediate
Spiral	Large	Circular	No	Yes
Paraboloid	Large	Linear or circular	Yes	High
Waveguide slots	Narrow	Linear	Yes	High
Lens	Large	Linear or circular	Yes	High

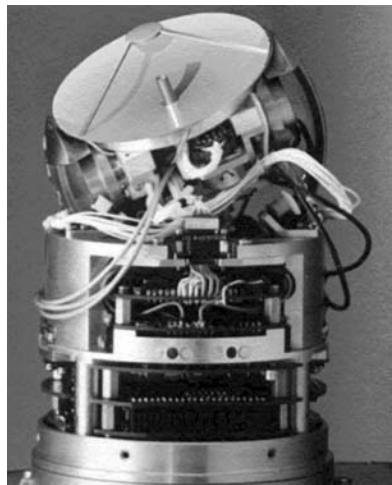
element types; virtually all of these can be implemented in a body-fixed configuration. However, mounting the antenna on a mechanical gimbal is more popular for most missile-guidance applications using proportional (or modern) guidance. A gimbaled slotted planar seeker antenna is shown in Figure 46-2. Detailed performance and design criteria for the basic antenna radiator types are found in the appropriate chapters of this Handbook.

Many system factors also must be considered in the antenna-design process. Antenna beamwidth determines whether spatial resolution of multiple targets occurs.⁵ The sooner that resolution takes place on a flight path, the more time the missile will have to correct errors induced by tracking the multiple-target centroid. Low sidelobes are important because they decrease electronic-jamming effects. This is particularly important if the missile gets significantly close to the jammer during the engagement. Noise injected in this manner can increase the final miss distance.

In a monopulse tracking system, low sidelobes are necessary in both the sum and difference modes in order to reduce the antenna's susceptibility to jamming threats. This creates the problem of achieving simultaneously low sidelobes in the sum and difference radiation patterns.⁴ Also, the selection and design of a monopulse feed network cannot proceed without consideration of errors and their effects on pattern performance. Precomparator phase shifts are more critical than postcomparator errors. After the comparator, any phase shifts or amplitude imbalances do not affect the position of the null but will decrease the sensitivity of the detector.⁶

The body-fixed antenna configuration confronts designers with unique problems. The beam may be very broad for a wide field of view (FOV), or it may be rapidly steered to form a tracking beam. The wide-FOV approach requires that the airframe and the autopilot be more restricted in their responsiveness. The steered-beam approach requires rapid beam forming and signal processing to isolate the missile-body rotational motion properly from the actual target motions. Some body-fixed antennas need to be mounted conformally to preserve the central forward viewing area for a second seeker in a dual-mode configuration.

It is often necessary to be able to integrate and test key receiver front-end elements and the RF elements of the transmitter as part of the seeker-antenna assembly. For example, it is generally possible to integrate in one package the feed antenna, monopulse comparator, mixer or mixers, local oscillator, and transmitter oscillator.⁷

**FIGURE 46-2** Antenna and gimbal assembly
(Courtesy of ST Technology)

46.3 IMPACT OF AIRFRAME ON ANTENNA DESIGN

The airframe and the autopilot steer the missile according to the guidance-computer commands to minimize the final miss distance at impact. The dynamic behavior of the missile in response to commands consists of attitude changes that cause the gimbal-mounted antenna to receive through an ever-changing portion of the radome.⁸ This, coupled with the natural geometry changes that occur as an engagement takes place, makes the antenna appear to have time-varying directivity and gain properties. The response characteristics of the airframe and the autopilot generate a pseudo-noise influence of the radome-antenna combination.

A definition of the angles relating to the seeker-radome and missile-target geometry is illustrated in Figure 46-3a. Here, the seeker antenna is shown, mounted on a mechanical gimbal whose pointing function is generally controlled by a servo loop. The difference between the actual missile-to-target line of sight (LOS) and the apparent LOS shown in this figure is the radome boresight error (BSE). Missile flight angles are referenced to airframe gyros, while antenna pointing angles are referenced to rate gyros mounted on the antenna gimbal.

For a body-fixed on-board sensor antenna, there is no track loop and the body isolation function is not present; the airframe gyros (roll, pitch, yaw) provide angular references for guidance toward the target. The implementations of these functions vary from system to system. Factors such as operational altitude and range, targets to be engaged, missile speed, and terrain over which missions are to be performed also influence the design.^{9,10}

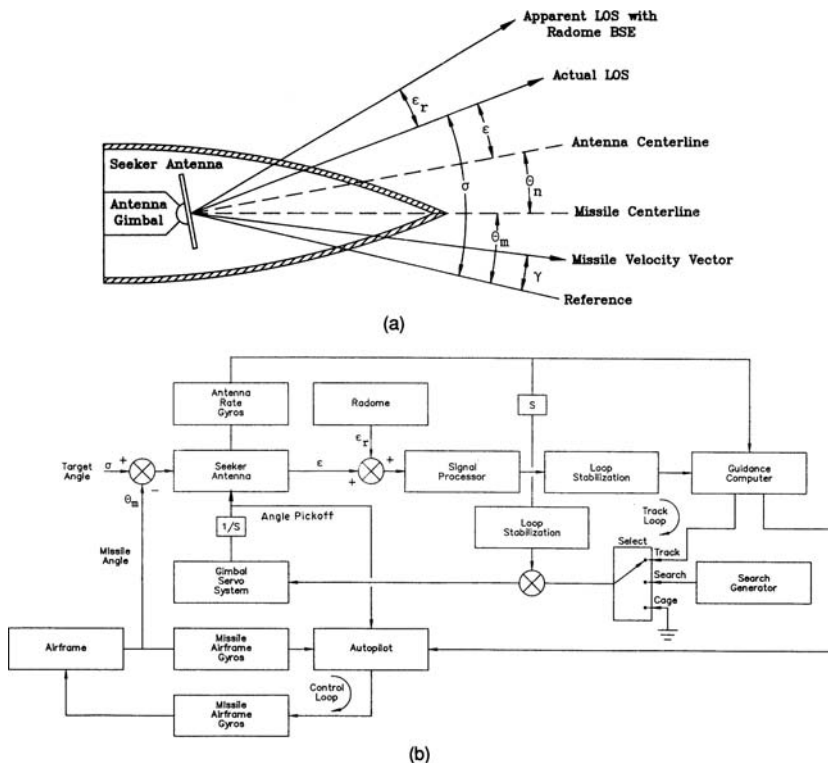


FIGURE 46-3 Seeker antenna and overall guidance functional diagram: (a) seeker/missile/target angular relationships and (b) tracking system block diagram

The missile-borne gimbaled guidance antenna enters the picture as depicted in the block diagram shown in Figure 46-3*b*. This figure shows the major functional features present in missiles and projectiles with on-board guidance. Through the track loop, the antenna gimbal platform is inertially stabilized in space relative to the LOS, and the monopulse error voltage drives a feedback loop to keep the antenna pointed toward the target.

The airframe and the autopilot steer the missile according to the guidance-computer commands to minimize the final miss distance at impact. The dynamic behavior of the missile in response to commands consists of attitude changes that cause the antenna to receive through an ever-changing portion of the radome. This, coupled with the natural geometry changes that occur as an engagement takes place, makes the antenna appear to have time-varying directivity and gain properties. The response characteristics of the airframe and the autopilot generate a pseudo-noise influence of the radome-antenna combination.

46.4 SEEKER-ANTENNA MECHANICAL CONSIDERATIONS

Generally, the seeker antenna is of the maximum size possible within the limits of the missile-body diameter in order to maximize antenna gain and minimize the antenna beamwidth. Alternatively, conformal- and flush-mounted antennas (most often forward mounted) can utilize the entire conical forward area to achieve maximum gain.¹¹

Gimbaled antenna systems often must mechanically steer not only the antenna structure but also the attendant beam-forming network and critical transmitter and receiver elements. As the mechanical-steering rates become excessive, the conformal-mounted antenna with electronically steered beams must be used.¹⁰⁻¹⁶ When all the seeker-antenna and associated hardware are gimbaled together, the use of lightweight materials may achieve the desired results. Lightweight techniques include the following:^{7,17}

- The use of lightweight honeycomb materials (including low-density dielectric foams) for both filler and structural-load-bearing surfaces. Foam reflectors and waveguide elements including feed horns can be machined and then metallized by vacuum deposition, plating, or similar techniques.
- The use of stripline techniques. Using multilayer techniques, one can have a layer with microstrip radiating elements, a layer with beam-forming networks, and a layer with beam-switching and signal-control elements such as a diode attenuator and phase shifters. The use of plated-through holes and/or pins to make RF connections can eliminate cables and connectors. Excess substrate can be removed and printed-circuit-board edges plated in lieu of the use of mode-suppression screws.
- The use of solid dielectric lenses, above Ku-band, with no significant weight penalty. Lenses can also be zoned to remove excess material.¹⁸

46.5 APERTURE TECHNIQUES FOR SEEKER ANTENNAS

Precise control of both amplitude and phase to permit aperture illumination tailoring is necessary for achieving a well-behaved antenna pattern.¹⁹⁻²⁵ Aperture tapering of rotationally symmetric illuminations starts with the feed itself and continues by varying the energy across the aperture by several methods: corporate-power splitting, aperture-element proportional size or spacing, surface control of the reflector or lens surface, and element thinning.

Circular antennas form the largest single class of seeker aperture antennas. The antenna patterns for circular aperture distributions that apply to seeker antennas are shown in other sections of the Handbook and in Kozakoff.²⁶ The case of cosine on a pedestal with $p = 0.3$ (–10.46 dB) edge illumination or $p = 0.2$ (–13.98 dB) edge illumination is practical when using horn feeds. The uniform illumination condition can be approximated using array techniques, but a tapered illumination is often preferred to achieve lower sidelobes.

Reflector or Lens Antennas

A consideration in the design of monopulse reflector antennas is the four-horn-feed design and the effects of aperture blockage. One criterion posed for an optimum monopulse feed configuration is maximizing the product of the sum times the derivative of the difference error signal.²⁷ The monopulse horn size for optimum monopulse sensitivity is shown in Figure 46-4. The impact of optimizing monopulse feed design on antenna performance is illustrated in Figure 46-5, in which a 5.5-in- (139.7-mm-) diameter aperture is assumed in the calculations. These data illustrate that at millimeter wavelengths (above 30 GHz), the effects of aperture blockage on antenna performance are small for this size of antenna even when a reflector is used.

Other criteria, such as the requirement for equal E- and H-plane antenna beamwidths, may enter into the details of monopulse feed design. An example of a monopulse-feed antenna operating near 95 GHz that meets equal-beamwidth requirements appears in Kozakoff and Britt.²⁸ For operation near 95 GHz, the individual waveguide apertures of the four horn feed were 0.080 by 0.100-in- (2.032- by 2.54-mm). They were fabricated using an electroforming process.

A wide-angle-scan capability can be achieved by using a twist-reflector concept with a rotatable planar mirror.^{29,30} In this configuration, the forward nose of the missile can be approximately paraboloidal-shaped, or the paraboloid can be located inside a higher-fineness-ratio radome. The paraboloid is composed of horizontal metal strips. The movable planar twist reflector uses 45°-oriented strips $1/4\lambda$ above the planar metallic reflector. (Grids may also be used to combine widely separated frequencies to permit dual-band operation.³¹)

Beam steering is obtained by moving the planar mirror. The steering is enhanced because for every 1° that the mirror moves, the beam moves 2°. By using a parallelogram gimbal to move the planar mirror, beam accelerations of $20,000^\circ/\text{s}^2$ have been demonstrated.³⁰

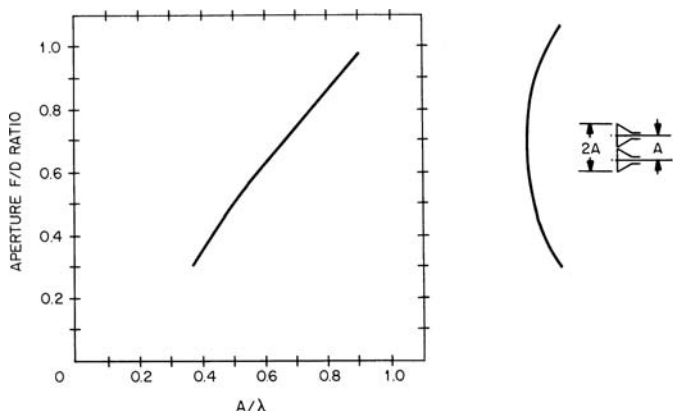


FIGURE 46-4 Monopulse antenna horn size or spacing for optimum monopulse sensitivity

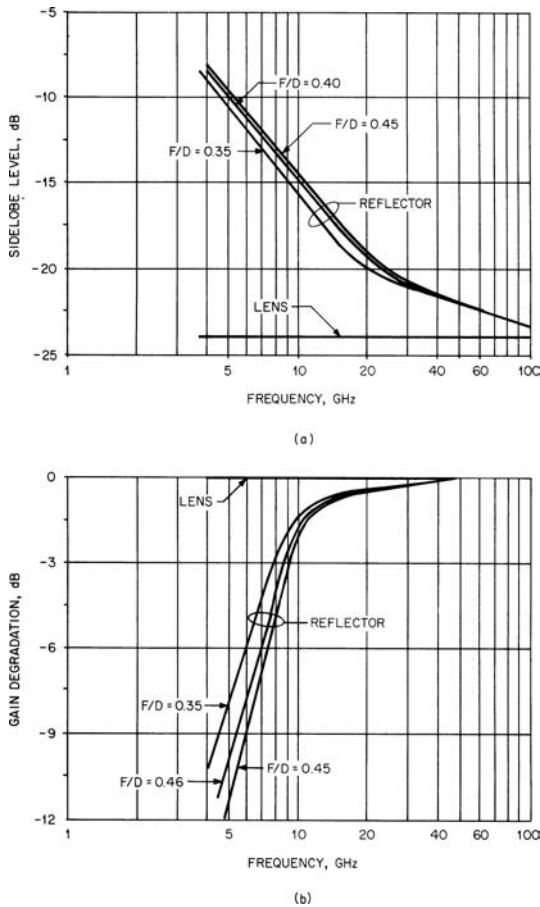


FIGURE 46-5 Performance of optimized four-horn monopulse-fed 5.5-in- (139.7-mm-) diameter apertures: (a) sidelobe level and (b) gain degradation

In this situation, a 42° beam scan was achieved with 21° of mirror motion, and less than 2 dB of sidelobe degradation of the sum pattern occurred.

By using a corrugated horn with a rotationally symmetric feed pattern providing a -17 -dB edge taper (capable of achieving a -30 -dB sidelobe level), one can further taper by varying the inner- and outer-surface contour of a collimating lens to achieve a circular Taylor amplitude distribution ($\bar{n} = 7$) with a -40 -dB sidelobe level.¹⁸ Practically, one can make the lens out of a material having an $\epsilon_r = 6.45$ (titania-loaded polystyrene) and follow up with a surface-matching layer made out of a material having an $\epsilon_r = 2.54$ (Rexolite-polystyrene).

Ring-Array, Planar-Array, and Conformal Antennas

Array techniques can be used to provide a nearly planar (flat-plate) aperture. Commonly used are arrays that provide symmetric patterns and offer a maximum use of the available

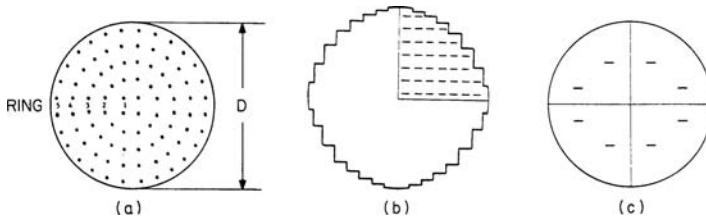


FIGURE 46-6 Monopulse planar-array geometries: (a) five-ring monopulse array, (b) waveguide slot array, and (c) thinned-slot (image) ring array

circular aperture with reduced grating-lobe potential. In such an array, higher aperture efficiency with low sidelobes can be achieved because of reduced aperture blockage and spill-over. In monopulse applications, it is noted that the ring array contains no central element and exhibits quadrantal symmetry, as indicated in Figure 46-6.

The choices in a ring-array design include the number of rings and the number of elements in each ring and the feed network.^{32,33} Constraints include the following:

- Minimum distance of the outer elements to the antenna edge
- Minimum spacing of the elements as impacted by excess mutual coupling (between elements and quadrants and feed geometry)
- Desired radiation pattern (as impacted by the density tapering)

Optimization of small ring arrays (having diameters $< 10\lambda$) is often carried out heuristically or nonlinearly. Examples of array geometries that have been analytically determined to have -17 - to -20 -dB sidelobes are given in Table 46-2 for the array geometry shown in Figure 46-6a.

Array elements can be of many types.³⁴⁻⁴¹ A two-dimensional array of waveguide slots excited by a network of parallel waveguides forming the antenna structural supports has been used.⁴⁰ Waveguide slot arrays have demonstrated low cross-polarized response as well. The bandwidth of a waveguide slot array is inversely proportional to the size of the array. A variety of outer contours can be utilized to conform to the space available.

Thinning techniques for these types of arrays reduce complexity.⁴² By using an image-element approach, a reduction of the number of elements by over 90 percent is possible (see Figure 46-6c). This method also readily allows for integration of the monopulse comparator.

TABLE 46-2 Data on Five-Ring Monopulse Uniformly Excited Array (Array OD = 8.16λ)³²

Ring number	Case A*	Case B*
	Directivity = 27.6 dB Maximum Sidelobe Level = -20 dB	Directivity = 28.1 dB Maximum sidelobe level = -17 dB
	Radius, λ	Radius, λ
r_1	0.45	0.50
r_2	1.25	1.33
r_3	2.08	2.16
r_4	3.06	2.98
r_5	3.72	3.72

*The maximum outer diameter = 8.16λ for both cases.

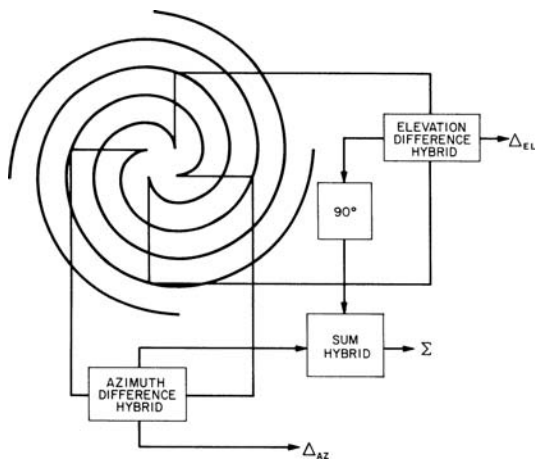


FIGURE 46-7 Monopulse circuit for a four-arm spiral antenna

Multimode and Single-Mode Spirals

The multimode spiral antenna is a very broadband, broad-beamed, circularly polarized antenna that is amenable to ARH (antiradiation homing) applications.^{43,44} Such an antenna can also be synthesized by using a circular array.⁴⁵ A simplified circuit to resolve monopulse sum-and-difference patterns is shown in Figure 46-7; typical angular coverage is from $\pm 30^\circ$ to $\pm 40^\circ$. The rather wide instantaneous FOV makes these antennas attractive for body-fixed applications.

The printed-circuit construction of four-arm spirals limits their use to low-power applications. Thus, this type of antenna is almost always used for passive (nontransmitting) applications. Loading techniques can be employed to reduce size, permitting operation at low frequencies. Current fabrication technology permits high-frequency operation to above 40 GHz. Single-mode spirals are also very useful as elements in small arrays. For operation between 2:1 and 3:1 bandwidths, small pyramidal dielectric elements or horns can be placed on the spiral front surface. The conical extensions act as endfire elements that narrow the element beamwidth and aid in suppressing array grating lobes.⁴⁶

46.6 PACKAGING SEEKER ANTENNAS

The seeker antenna functions require some unique considerations.

Conformal Array

Because the typical tactical-sized missile is about six inches in diameter, the seeker antenna must be packaged to accommodate both its own function and be compatible with other missile functions as well. Newer missiles are capable of combining two or more sensing techniques. Thus, clever arrangements of components are needed to accommodate the hardware involved in the seeker functions.

One antenna location, shown in Figure 46-8, is a cylindrical array covering the complete circumference that is located in the forward portion of the missile, offering the forward viewing area for another sensor. The array elements are essentially flush mounted with a minimal depth into the missile interior.

Assuming an actual 5.7-in (144.8-mm) diameter is available, the total array circumferential shaded length about 18-in-(457.2-mm-). This length could accommodate 18 elements of nominal 1-in (25.4-mm) circumferential width. The axial length can be longer than 1 in (25.4 mm). Microstrip antenna elements lend themselves to the seeker package. The array element mounting and dielectric radome cover are compatible with compact seeker electronics.

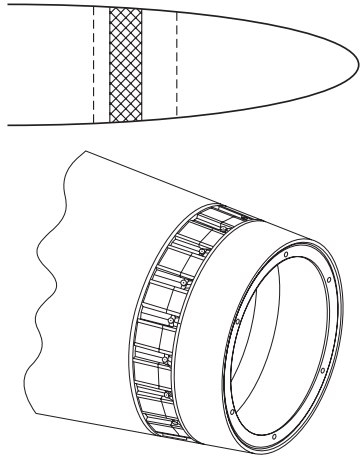


FIGURE 46-8 Conformal-mounted circumferential seeker antenna array. The portion of the missile forward of the array can be used for another seeker. (Adapted from U.S. patent 6,407,711)

Forward Viewing Array

The seeker antenna may need to provide a dual-polarized monopulse capability and be packaged to work intimately with another sensor. Figure 46-9 shows a similar 6.0-in- (152.4-mm-) diameter antenna package offering this capability. The antenna is comprised of four dual-polarized slotted waveguide subarrays arranged to form an aperture and combined to form the sum, difference, azimuth, and elevation signals, each with the correct sense of circular polarization.

The dual-polarized antenna design uses balanced orthogonal polarizations. This approach provides low cross-polarization interference performance. The use of the slot array provides low loss, compact size, and high-volume producibility.

The antenna is used for the millimeter wave portion of single- or multimode seekers. It is comprised of three major subsystems: a three-layer, meander-line linear-to-circular polarizer; a dual-linear waveguide slot array; and a dual-channel, waveguide monopulse beamforming network. A through-hole in the center of the antenna can accommodate other sensing modes such as laser and infrared.



FIGURE 46-9 Dual-polarized monopulse seeker antenna (Courtesy of EMS Technologies, Inc.)

46.7 SMALL ARRAYS FOR SEEKER ANTENNAS

Arrays of two, three, four, and five elements can be used for seeker antennas in small interferometers. These types of arrays are usually fixed-mounted (nongimballed) and used when pursuit navigation algorithms are suitable. With the addition of a gyroscope to sense missile angular motion, line-of-sight data can be derived.⁴⁷

Good sum-and-difference or phase monopulse angle-tracking performance can be achieved when the array elements, the array beamwidth, and the array element spacing are properly chosen. Broadband performance is often achievable by using antenna elements such as spirals or log-periodic antennas. Small arrays can be used by themselves, or they can illuminate apertures such as reflectors or lenses.

A four-element square or diamond array can be sized to fit within a given missile diameter using

$$D_{\text{missile}} = d_{\text{spiral}}(\sqrt{2} + 1)$$

The spiral diameter is sized for its lowest frequency of operation. Mode 1 spirals usually have a maximum diameter bounded by an upper value of $\lambda/2$ and a nominal value of λ/π . For operation to as low as 2 GHz, the desired missile diameter is between 7.12 in ($\lambda/2$ criterion) or 4.54 in (λ/π criterion). This condition occurs when the spirals of Figure 46-10 are enlarged so that they just touch and collectively fill the full missile diameter. The upper frequency of this type of array is limited by array pattern grating lobes.

Control of the array element pattern is important in reducing energy spillover beyond the difference-pattern lobes. As noted earlier, these lobes and the sum pattern grating lobes can be limited by adding dielectric rods in front of the spiral to cause the spiral antenna element pattern to narrow as the frequency increases. A 3:1 band can usually be achieved.

A four-element array can serve as an RF seeker by utilizing a phase processor. The effects of system errors for such an array on antenna patterns, S curves, and angular sensitivity as ascertained analytically can be used to form a catalog of results that can be employed to diagnose fabrication errors or tolerances based on measured data.⁴⁸

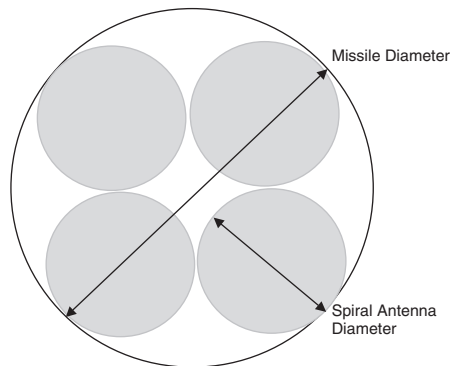


FIGURE 46-10 Four-element square or diamond array fitted within a given missile diameter

46.8 RADOME EFFECTS

The introduction of a protective radome over a seeker antenna results in an apparent boresight-error (BSE) because of the wavefront phase modification over the seeker antenna aperture. A published PC-based software radome code can be used to compute the electrical effects of a radome over a seeker antenna.⁴⁹ These effects need to be carefully considered in the selection of a radome design because they impact seeker-system performance.^{9,50-52}

In the case of pursuit guidance, the miss distance is proportional to the BSE. Conversely, in the case of proportional guidance, it is the boresight-error slope (BSES) that more directly impacts miss distance. The relationship between BSE or BSES and miss distance is not straightforward but generally is quantified only by a hardware-in-the-loop (HWIL)^{53,54} or computer simulation that takes into account a variety of scenarios and flight conditions. A BSE requirement of better than 10 mrad (or a BSES requirement of better than 0.05 deg/deg) and maximum transmission loss of 1.0 dB are typical for many radome types.

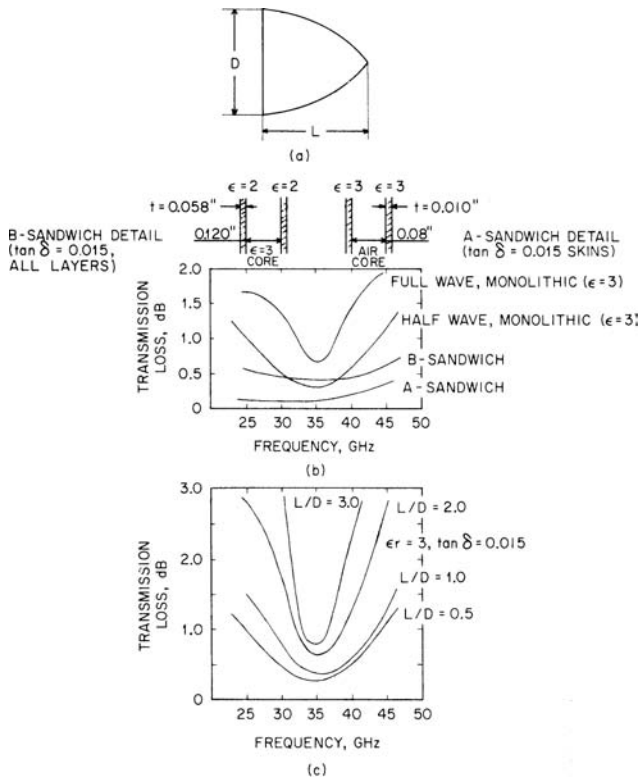


FIGURE 46-11 Missile-radome transmission data: (a) radome geometry, (b) hemispherical-radome ($L/D = 0.5$) performance for a low-velocity ($< \text{Mach } 1$) missile, and (c) monolithic half-wave radome performance for various fineness (L/D) ratios

The method most often employed for the evaluation of radome effects on seeker antenna performance utilizes open-loop testing in a suitable anechoic-chamber facility.⁵⁵ This requires a precalibration of the monopulse error-channel sensitivity in volts per degree without the radome and subsequently measuring the error-channel outputs with the radome over the seeker antenna.

Broadband ARH antennas generally require the use of multilayer radome walls to obtain a required radome bandwidth commensurate with antenna performance. Figure 46-11 illustrates hemispherical-radome performance for half-wave, full-wave, A, and B sandwich walls. Half-wave radomes generally have a useful bandwidth on the order of 5 to 20 percent, depending on dielectric constant and radome-fineness ratio. Higher-order wall radomes have considerably narrower useful bandwidths and thus are limited to narrowband-antenna applications. Figure 46-11 also demonstrates transmission loss versus frequency for a half-wave wall radome when radome-fineness ratio is taken as a parameter.

46.9 EVALUATION OF SEEKER ANTENNAS

Seeker-antenna testing proceeds from conventional methods utilizing anechoic-chamber techniques with the antenna alone.⁵⁶⁻⁶¹ Next, open-loop testing including critical elements of the seeker electronics (*S* curves) is most often performed. Ultimately, a variety of simulations are used to ascertain the seeker-antenna performance in both benign and complex environments.

The many nonlinear elements in the guided missile make complete closed-form analytical simulations unsuitable for final analyses. The precise simulation of these elements is also difficult. The final analyses are best done with an HWIL simulation.⁵²⁻⁵⁴ As many hardware items as possible should be inserted, leaving only the propulsion, aerodynamics, and instrument feedback to be simulated analytically.

There are several concepts for implementing a simulation with seeker hardware. Placing the seeker in an anechoic chamber with the target of appropriate characteristics moved about according to the engagement geometry is one method. Flying a target with the seeker on an orientation table is another. The HWIL simulation serves as a tool not only for statistical performance analysis but for provision of validity information for an improved all-analytical model of the system. The HWIL method provides the only valid present means of evaluating the interaction of the radome-antenna combination in a missile-guidance-seeker environment.

46.10 ADDITIONAL DEVELOPMENTS AND SOURCES OF INFORMATION

A number of allied technology factors⁶²⁻⁶³ that affect the design and fabrication of seeker antennas are summarized in the following:

- Low-mass antenna techniques for high-*g* environments:
 - Use of negative space, i.e., plated hollow areas for feeds and transmission lines, e.g., horns, waveguides
 - Cannon-launched projectile-mounted antennas must withstand *g*-forces at launch of 10,000 to 16,000 *g*
 - Use of physically small (but electrically large) apertures, including classic slots, dipoles, patches, spirals, and Vivaldi slots/notches
 - Use of thin, flat, printed-circuit reflector antennas⁶⁴
- Advanced seeker-antenna manufacturing techniques:
 - Higher millimeter frequencies, i.e., above 100 GHz
 - Nonplanar geometries, i.e., conical spirals
 - Slotted waveguide antennas at Ka-band and higher⁶⁵ and waveguide monopulse comparators at W-band
 - Millimeter-wave microstrip antennas (including annular rings) and microstrip comparators⁴
 - Better packaging techniques to divide aperture space among several antenna requirements^{2,66}

- Small four-element arrays using both amplitude and phase monopulse techniques (amplitude monopulse used to resolve phase ambiguities) and the use of GPS for tactical missile navigation.⁶⁷
- Use of the U.S. Patent Office (<http://patents.uspto.gov/>) to assess extensive information on seeker antenna concepts and fabrication techniques. A short list of related U.S. patents follows:
 - 6,407,711 Antenna Array Apparatus with Conformal Mounting Structure
 - 6,127,985 Dual-polarized Slotted Array Antenna
 - 5,818,393 Wide Field-of-View Fixed Body Conformal Antenna Direction Finding Array
 - 5,686,929 RF Homing Head Antenna System for Missiles
 - 5,483,894 Integral Missile Antenna Fuselage Assembly
 - 5,344,099 Missile Beamrider Guidance Using Polarization Agile Beams
 - 5,307,077 Multi-spectral Seeker Antenna
 - 5,218,361 Interferometer Ambiguity Resolution Using Missile Roll
 - 4,577,196 Missile Mounted Waveguide Antenna
 - 4,431,996 Missile Multi-frequency Antenna
 - 4,204,212 Conformal Spiral Antenna
 - 4,194,204 High Resolution Microwave Seeker

REFERENCES

1. M. I. Skolnik, *Radar Handbook* (New York: McGraw-Hill, 1978).
2. R. A. Sparks, "Systems Applications of Mechanically Scanned Slotted Array Antennas," *Microwave J.*, vol. 31, no. 6 (June 1988): 26–48.
3. J. B. Mead and R. E. McIntosh, "Polarimetric Remote Sensing," *PIER 3, Progress in Electromagnetics Research*, J. A. Kong (ed.) (New York: Elsevier Science Publishing, 1990): 391–450.
4. F. Lalezari, T. C. Boone, and J. M. Rogers, "Planar Millimeter-Wave Arrays," *Microwave J.*, vol. 34, no. 4 (April 1991): 85–92.
5. H. G. Oltman and M. E. Beebe, "Millimeter Wave Seeker Technology," *AIAA Guid. Cont. Conf.* (August 1978): 148–158.
6. C. K. Watson and K. Ringer, "Feed Network Design for Airborne Monopulse Slot-Array Antennas," *Microwave J.* (June 1988): 129–145.
7. C. R. Seashore et al, "MM-Wave Radar and Radiometer Sensors for Guidance Systems," *Microwave J.* (August 1979): 41–59.
8. H. L. Hirsch and D. C. Grove, *Practical Simulation of Radar Antennas and Radomes* (Norwood, MA: Artech House, 1987).
9. P. Garnell and D. J. East, *Guided Weapons Control Systems* (London: Pergamon Press, 1977).
10. R. A. Chervolk, "Coherent Active Seeker Guidance Concepts for Tactical Missiles," *IEEE Eascon Rec.* (September 1978): 199–202.
11. R. C. Hansen (ed.), *Conformal Antenna Array Design Handbook*, AIR310E (Washington: U.S. Navy, September 1981).
12. P. C. Bargeliotis, A. F. Seaton, A. T. Villeneuve, and W.H. Kummer, "Conformal Phased Array Breadboard," Final Rep. on NOO19-76-C-0495, ADA038-350, January 1977.
13. T. W. Bazire et al, "A Printed Antenna/Radome Assembly (RADANT) for Airborne Doppler Navigation Radar," *Fourth Europ. Microwave Conf. Proc.* (September 1974): 494–498.

14. H. S. Jones, Jr., "Some Novel Design Techniques for Conformal Antennas," *IEE Conf. Antennas Propagat.* (November 28–30, 1978): 448–452.
15. Naval Air Systems Command, "Conformal Antennas Research Program Review and Workshop," AD-A015-630, April 1975.
16. A. T. Villeneuve et al., "Wide-Angle Scanning of Linear Arrays Located on Cones," *IEEE Trans. on Antennas and Propagat.*, vol. AP-22 (January 1974): 97–103.
17. J. S. Yee and W. J. Furlong, "An Extremely Lightweight Electronically Steerable Microstrip Phased Array Antenna," *IEEE Antennas Propagat. Int. Symp. Dig.* (May 1978): 170–173.
18. D. K. Waiteo, "Lens Design for Arbitrary Aperture Illumination," *IEEE Antennas Propagat. Int. Symp. Dig.* (October 1976): 476–479.
19. E. T. Bayliss, "Design of Monopulse Antenna Difference Patterns with Low Sidelobes," *Bell Syst. Tech. J.* (May–June 1968): 623–650.
20. L. J. Du and D. J. Scheer, "Microwave Lens Design for a Conical Horn Antenna," *Microwave J.* (September 1976): 49–52.
21. P. W. Hannan, "Optimum Feeds for All Three Modes of a Monopulse Antenna (Parts I and II)," *IRE Trans. Antennas Propagat.*, vol. AP-9 (September 1961): 444–464.
22. R. W. Kreutel, "Off-Axis Characteristics of the Hyperboloidal Lens Antenna," *IEEE Antennas Propagat. Int. Symp. Dig.* (May 1978): 231–234.
23. S. Pizette and J. Toth, "Monopulse Networks for a Multielement Feed with Independent Control of the Three Monopulse Modes," *IEEE Microwave Theory Tech. Int. Symp. Dig.* (April–May 1979): 456–458.
24. R. L. Sak and A. Sarremejean, "The Performance of the Square and Conical Horns as Monopulse Feeds in the Millimetric Band," *IEE Ninth Europ. Microwave Conf. Proc.* (September 1979): 191–195.
25. P. A. Watson and S. I. Ghobrial, "Off-Axis Polarization Characteristics of Cassegrainian and Front-Fed Antennas," *IEEE Trans. on Antennas and Propagat.*, vol. AP-20 (November 1972): 691–698.
26. D. J. Kozakoff, "Aperture Antennas," *Wiley Encyclopedia of Electrical and Electronics* (Madison, WI: John Wiley and Sons, Inc., 1998).
27. D. R. Rhodes, *Introduction to Monopulse* (New York: McGraw-Hill, 1959).
28. D. J. Kozakoff and P. P. Britt, "A 94.5 GHz Variable Beamwidth Zoned Lens Monopulse Antenna," *IEEE Southeastcon Proc.* (April 1980): 65–68.
29. E. O. Houseman, Jr., "A Millimeter Wave Polarization Twist Antenna," *IEEE Antennas Propagat. Int. Symp. Dig.* (May 1978): 51–53.
30. D. K. Waiteo and J. F. Koneczny, "Millimeter Wave Monopulse Antenna with Rapid Scan Capability," *IEEE Antennas Propagat. Int. Symp. Dig.* (June 1979): 477–480.
31. L. Goldstone, "Dual Frequency Antenna Locates MM Wave Transmitters," *Microwave Syst. N.* (October 1981): 69–74.
32. D. A. Huebner, "Design and Optimization of Small Concentric Ring Arrays," *IEEE Antennas Propagat. Int. Symp. Dig.* (May 1978): 455–458.
33. A. R. Lopez, "Monopulse Networks for Series-Feeding an Array Antenna," *IEEE Trans. on Antennas and Propagat.*, vol. AP-16 (July 1968): 68–71.
34. S. W. Bartley and D. A. Huebner, "A Dual Beam Low Sidelobe Microstrip Array," *IEEE Antennas Propagat. Int. Symp. Dig.* (June 1979): 130–133.
35. C. W. Garven et al., "Missile Base Mounted Microstrip Antennas," *IEEE Trans. on Antennas and Propagat.*, vol. AP-25 (September 1977): 604, 616.
36. C. S. Malagisi, "Microstrip Disc Element Reflect Array," *IEEE Eascon Rec.* (September 1978): 186–192.
37. S. Nishimura et al., "Franklin-Type Microstrip Line Antenna," *IEEE Antennas Propagat. Int. Symp. Dig.* (June 1979): 134–137.
38. P. K. Park and R. S. Elliott, "Design of Collinear Longitudinal Slot Arrays Fed by Boxed Stripline," *IEEE Trans. on Antennas and Propagat.*, vol. AP-29 (January 1981): 135–140.
39. Flat Plate Antennas, Rantec Div., Emerson Electric Co., June 1969.

40. G. G. Sanford and L. Klein, "Increasing the Beamwidth of a Microstrip Radiating Element," *IEEE Antennas Propagat. Int. Symp. Dig.* (June 1979): 126–129.
41. D. H. Schaubert et al., "Microstrip Antennas with Frequency Agility and Polarization Diversity," *IEEE Trans. Antennas Propagat.*, vol. AP-29 (January 1981): 118–123.
42. W. H. Sasser, "A Highly Thinned Array Using the Image Element," *IEEE Antennas Propagat. Int. Symp. Dig.* (June 1980): 150–153.
43. J. D. Dyson, "Multimode Logarithmic Spiral Antennas," *Nat. Electron. Conf.*, vol. 17 (October 1961): 206–213.
44. R. G. Corzine and J. A. Mosko, *Four-Arm Spiral Antennas* (Norwood, MA: Artech House, 1990).
45. J. M. Schuchardt and W. O. Purcell, "A Broadband Direction Finding Receiving System," *Martin Marietta Interdiv. Antenna Symp.* (August 1967): 1–14.
46. T. E. Morgan, "Reduced Size Spiral Antenna," *IEE Ninth Europ. Microwave Conf. Proc.* (September 1979): 181–185.
47. E. R. Feagler, "The Interferometer as a Sensor for Missile Guidance," *IEEE Eascon Rec.* (September 1978): 203–210.
48. L. L. Webb, "Analysis of Field-of-View versus Accuracy for a Microwave Monopulse," *IEEE Southeascon Proc.* (April 1973): 63–66.
49. D. J. Kozakoff, *Analysis of Radome Enclosed Antennas* (Norwood, MA: Artech House, 1997).
50. G. Marales, "Simulation of Electrical Design of Streamlined Radomes," *AIAA Summer Computer Simulation Conf.* (July 1979): 353–354.
51. A. Ossin et al., "Millimeter Wavelength Radomes," Rep. AFML-TR-79-4076, Wright-Patterson Air Force Base, Dayton, Ohio, July 1979.
52. K. Siwak et al., "Boresight Errors Induced by Missile Radomes," *IEEE Trans. on Antennas and Propagat.*, vol. AP-27 (November 1979): 832–841.
53. R. F. Russell and S. Massey, "Radio Frequency System Simulator," *AIAA Guidance Cont. Conf.* (August 1972): 72–861.
54. D. W. Sutherlin and C. L. Phillips, "Hardware-in-the-Loop Simulation of Antiradiation Missiles," *IEEE Southeastcon Proc.* (April 1976): 43–45.
55. J. M. Schuchardt et al., "Automated Radome Performance Evaluation in the RFSS Facility at MICOM," Proc. 15th EM Windows Symp., Atlanta, June 1980.
56. R. C. Hansen, "Effect of Field Amplitude Taper on Measured Antenna Gain and Sidelobes," *Electron. Lett.* (April 1981): 12–13.
57. R. C. Hansen et al., "Sidewall Induced Boresight Error in an Anechoic Chamber," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-7 (November 1971): 1211–1213.
58. D. J. Kaplan et al., "Rapid Planar Near Field Measurements," *Microwave J.* (January 1979): 75–77.
59. A. S. Thompson, "Boresight Shift in Phase Sensing Monopulse Antennas Due to Reflected Signals," *Microwave J.* (May 1966): 47–48.
60. R. D. Monroe and P. C. Gregory, "Missile Radar Guidance Laboratory," Range Instrumentation-Weapons System Testing and Related Techniques, AGARD-AG-219, vol. 219, 1976.
61. A. F. Ali, "Electronic Warfare Testing at the Benefield Anechoic Facility," *IEEE 1997 AUTOTESTCON Proceedings* (1997): 232–243.
62. D. A. James, "Seeker Antennas," *Radar Homing Guidance for Tactical Missiles* (London: Macmillan, Ltd., 1986): 38–55.
63. A. Henderson and J. R. James, "A Survey of Millimeter Wavelength Planar Antennas for Military Applications," *Radio Elect. Eng.*, vol. 52, no. 11/12 (1982): 543–580.
64. B. Huder and W. Menzel, "Flat Printed Reflector Antenna for MM-Wave Applications," *Electronics Lett.*, vol. 24, no. 6 (March 1988): 318–319.
65. H. P. Muhs, "MM-Wave Antenna," *Microwave J.*, vol. 28, no. 7 (July 1985): 191–194.
66. D. C. Collier, "Flat Plate Antennas," *Appl. Microwave* (Summer 1991): 117–123.
67. D. N. Pittman and C. E. Roberts, "Determining GPS Anti-Jamming Performance on Tactical Missiles," *IEEE Position Location and Navigation Symposium* (1994): 641–648.

Chapter 47

Direction Finding Antennas and Systems

Robert L. Kellogg
Eldon E. Mack
Cathy D. Crews

Argon ST, Fairfax, Virginia

CONTENTS

47.1	INTRODUCTION	47-2
47.2	DIRECTION FINDING SYSTEM DESIGN	47-3
47.3	DIRECTION FINDING SYSTEM EXAMPLES	47-8
47.4	DIRECTION FINDING ALGORITHMS	47-17
47.5	GEOLOCATION ALGORITHMS	47-26

47.1 INTRODUCTION

Although radio direction finding (DF) is as old as radio itself, new algorithms and new techniques are used in modern digital systems. This chapter looks at the classical underpinnings of DF and its modern applications.

Purpose of Direction Finding

In essence, DF is any technique that determines the direction of a radio transmitter.¹ More formally, radio DF is defined by the National Telecommunications and Information Administration (NTIA) as: “the determination of the position, velocity and/or other characteristics of an object, or the obtaining of information relating to these parameters, by means of the propagation properties of radio waves.”² Radio DF systems as well as radar and navigation systems are registered in the U.S. Patent system under Class 342.

Application of Direction Finding Systems

The practical applications of DF are numerous:

- **Radio navigation and satellite global positioning services** The terrestrial Long Range Navigation (LORAN) uses phase-comparison of beacon signals for hyperbolic line of bearing (LOB) and geolocation. The U.S. satellite Global Positioning System (GPS) and the Russian counterpart Global Navigation Satellite System (GLONASS) provide time of arrival signals for measuring precise geolocation and velocity.
- **Search and rescue services** The U.S. Coast Guard and worldwide organizations such as COSPAS-SARSAT monitor emergency radio channels for rescue. Both terrestrial and satellite DF networks are used to continuously locate emergency signals.
- **Signal direction finding and location systems** These systems operate throughout the radio spectrum from the high-frequency (HF) band to 18 GHz for purposes that range from animal tracking to locating stolen cars to gathering military intelligence.
- **Homing systems** These are typically small DF systems designed to guide a weapon or artillery shell toward a location. Active RF systems use radar, while passive systems can DF target emitter signals (i.e., “home on jam”).
- **Warning systems** Also called Electronic Support Measure (ESM) systems. These are typically rapid response DF and identification systems to protect against hostile radars and homing systems.
- **Radio astronomy** These are characterized by large antenna dishes or arrays that can be synchronized into large interferometric systems spanning continents. The resulting radio maps show density and velocity of natural signals from spin-flip states of hydrogen and various molecular energy state transitions.
- **Smart antennas** Recent advances in adaptive arrays or “smart antennas” provide spatial isolation of emitters, interference cancellation, and resistance to multipath signal fading³ (see Chapter 25).

Brief History of Direction Finding

Heinrich Hertz began radio DF with experiments in the directionality of antennas in 1888. The first mobile DF system occurred in 1906 when the Stone Radio & Telegraph Co.

installed a direction-indicating device on the U.S. naval collier *Lebanon*. The device was not particularly effective as it “necessitated the swinging of the ship to obtain the maximum signal [strength] and the bearing.”⁴ In 1915 Dr. Frederick A. Kolster, a former Stone Radio & Telegraph employee working at the U.S. Bureau of Standards, discovered that wire wound on a rectangular frame could be rotated to determine the direction of a signal. One year later, 20 Kolster loop DF systems called the “SE 74” were installed on U.S. ships.⁵

Radio navigation service experiments continued at the U.S. Bureau of Standards, resulting in the first workable aircraft DF system in the early 1920s.⁶ The first aircraft DF antennas were manually rotated, but by the 1940s, systems such as the Japanese Type 1-3 system and the Fairchild Aero Compass from which it was copied, provided a twin-loop antenna rotated by a motor.⁷

In the mid-1920s Robert Watson-Watt, inventor of radar, conducted DF research at Ditton Park. Co-researcher Edward Appleton rediscovered the Heavyside propagation layer 300 km above the earth’s atmosphere, which Watson-Watt dubbed the “ionosphere.”⁸ By 1937 the U.S. Navy was using the ionosphere for long-range High Frequency Direction Finding (HFDF, pronounced “Huff-Duff” by old-timers) and in 1941 established both the Atlantic and Mid-Pacific Strategic Direction-Finder Nets, giving Allies the worldwide capability to track German submarines.⁹ In 1943 the Germans improved HFDF with the Wullenweber circularly disposed antenna array (CDAA). After the war, it was copied by the Russians (naming their version “Krug”) and by the U.S. (whose operators called it the “Elephant Cage” for its size).

The LORAN system was also an outcome of World War II. Developed at the Massachusetts Institute of Technology, the system used a set of continuous wave beacons that could be phase differenced for hyperbolic contours of constant phase. LORAN-C replaced the original LORAN in the 1950s using pulsed signals, resulting in 0.5-km accuracy.

The mathematics for using Doppler direction finding were first developed by the Russian Kotelnikov,¹⁰ but it was the U.S. who launched the first operational satellite system in 1960. The system was called “Transit” and it used a cluster of six low earth orbit (LEO) satellites to provide precise 150/400-MHz reference transmissions.¹¹ In 1967 the Soviet Union launched Tsyklon, a similar Doppler navigation system.¹² Today the U.S. NAVSTAR GPS and its Russian counterpart, GLONASS, provide precise time, position, and velocity determination worldwide.

Using the same Doppler method as for navigation, Lehan and Brown received U.S. Patent 3063048 for a satellite search and rescue location system.¹³ In 1977 the COSPAS-SARSAT satellite search and rescue system was realized as a joint U.S.-Soviet effort that now includes over 35 countries from Algeria to Vietnam.¹⁴

Since 1980 satellite and terrestrial DF systems have gone through a digital revolution, replacing analog DF processing with digital algorithms such as Bartlett correlation and Schmidt’s now famous Multiple Signal Classification (MUSIC) algorithm.¹⁵

Section 47.2 examines DF systems and sources of error. Section 47.3 looks at specific antenna array systems, while Section 47.4 surveys digital DF algorithms. Section 47.5 explains two techniques for combining LOBs into geolocations, and because of the parallel in approach, Section 47.5 presents a basic Time Difference of Arrival (TDOA) geolocation algorithm.

47.2 DIRECTION FINDING SYSTEM DESIGN

DF systems with antenna arrays determine a signal’s LOB by using antenna voltage measurements. While signal direction may be sufficient for many applications, a network of DF systems can combine LOBs to form signal “fixes” or geolocations. Systems that measure time and frequency without antenna arrays may work in a Time Difference of Arrival (TDOA) network to determine signal geolocation directly without LOBs.

System Components

At its core, a DF system requires an antenna array, a receiver to amplify and filter antenna radio frequency (RF) voltages, some means to process the antenna array voltages into a signal angle of arrival (AOA), and an operator display (see Figure 47-1, solid boxes). Modern DF systems now include automated signal detection, classification, digital processing, and remote display (see Figure 47-1, dotted boxes). In all, modern DF systems are comprised of the following elements (numbers correspond to those in Figure 47-1):

1. An antenna or antenna array to collect signal energy from an emitted signal. Arrays can span continents, be compact in a handheld wireless device, or be virtual extensions of an existing array, exploiting techniques such as ESPRIT.¹⁶
2. An RF distribution system to provide interference filtering, antenna element combining, antenna element selection or switching, and distribution of RF output to the rest of the DF system. An important attribute of the RF distribution system is to allow calibration signals to be switched or multiplexed with the RF input.
3. A calibration system that provides known noise characteristics or test signal amplitude and phase to allow measurement of filter, cable, and receiver effects on system estimates of time, amplitude, and phase.
4. A receiving system to measure the response of the antenna system. RF radio waves are transformed into electron currents. Historic systems used analog measurements and processes to determine DF. Modern systems use analog to digital (A/D) converters to represent signal measurements in digital form.
5. A time and frequency reference system. Modern systems that digitize voltages and use both amplitude and phase information need all receivers and A/D converters synchronized.
6. A DF processor to extract radio determination information such as signal LOB and elevation angle. DF processors may estimate geolocation from multiple signals (e.g., navigation satellites), from coordinated DF systems (e.g. a worldwide HFDF network), or from a single DF system using running LOBs.
7. The DF processor may need extensive antenna array manifold calibration data on the response of the array at different signal angles and different frequencies. Some DF processors, such as the Watson-Watt DF system, use a predetermined mathematical formulation of the signal response. Correlation and eigen decomposition systems process the signal covariance matrix.
8. A signal detection and qualification processor. It is important for the DF processor to provide valid measurement output on signal energy. Further, some systems may qualify the energy as a “signal of interest” before performing DF.
9. In modern DF systems, there is usually a function of signal parameter collection, including time, frequency, angle of arrival, signal strength, etc.
10. If the DF system is on a mobile platform, platform position, velocity, time (PVT), attitude, and heading reference system (AHRS) information is required. This allows transformation of platform measured AOA to earth fixed LOB.
11. Network infrastructure allows control and display to be distributed worldwide to a multitude of users.
12. What was once a simple cockpit indicator has evolved into remote displays of LOBs, emitter location estimates, and confidence error ellipses on a geographical information system (GIS) used by one or many operators.

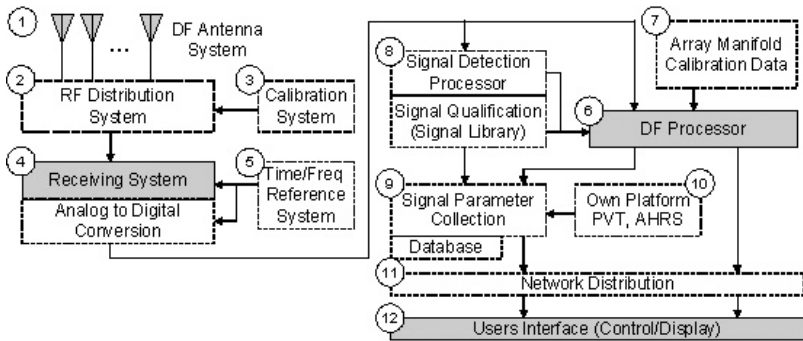


FIGURE 47-1 Traditional (solid boxes) and modern (dotted boxes) DF system design

The use of real-time signal-processing techniques now makes it possible to use phase and amplitude responses, employ antennas of any polarization, and to dynamically calibrate the receiver system for accurate measurements. Now the Internet makes it possible for remote command and control and worldwide data distribution.

System Accuracy and Sources of Error

Direction finding and geolocation system performance are measured primarily in terms of accuracy, timeliness, and throughput. DF accuracy is usually specified in degrees root mean squared (rms), while TDOA systems use rms time. Geolocation accuracy for both is described in terms of a confidence containment ellipse and occasionally in terms of Geometric Dilution of Precision (GDOP). Emitter to receiving site geometry is important and can greatly influence geolocation accuracy. The best geolocation accuracy is obtained when the emitter is in the center of an equilateral triangle of sensor sites.

For DF AOA systems, as a rule of thumb, the array DF accuracy is inversely proportional to the array aperture size. The upper frequency limit of the array is set by the Nyquist spacing of array elements. Ambiguous AOA occurs if the spacing exceeds $c/2f$ where c is the speed of light and f is the highest frequency of operation. DF accuracy degrades proportional to $1/f$ (see Figure 47-2), and there is no distinct bound for lowest frequency of operation. Most common wideband antenna arrays span a frequency range of 1:5 up to 1:8. For example, the VHF/UHF range can be broken into arrays that span 100–500 MHz and 500–3000 MHz.

An array can be made “sparse” when antenna elements are left out of a periodic structure. For example, a simple uniform linear array can be altered to minimal redundancy¹⁷ by leaving certain elements out. Additional elements can be added to an array to provide a larger eigen processing space that gives interference cancellation capability, immunity to fading, and increased hearability, especially in the presence of co-channel signals.

Other key parameters influencing DF accuracy include signal-to-noise ratio (SNR) resulting from a combination of emitter power, propagation distance and path mode (e.g. HF ground wave, ordinary and extraordinary skywave paths, and ionospheric tilting), atmospheric propagation (ducting, especially at VHF and UHF frequencies), atmospheric noise (e.g. lightning in HF), manmade noise (e.g. power line noise, automobile, broadcast

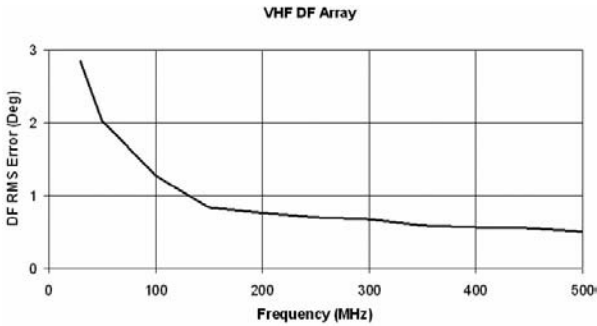


FIGURE 47-2 DF accuracy vs. frequency for VHF array 100–500 MHz

transmitters, etc.), and system internals (antenna gain, cable loss, receiver noise figure, and signal integration time). All of these effects can degrade and possibly corrupt DF results.

Antenna arrays and the systems behind them are usually calibrated and characterized under high SNR conditions, and whenever possible under threshold or minimum SNR conditions. The high SNR conditions establish instrumentation error, while threshold SNR testing more closely predicts operational performance. DF rms error and the Cramer-Rao bound is discussed at the end of Section 47.4.

For TDOA systems, Stein¹⁸ shows that the rms error in measurement from cross correlating two signal samples is based on the Cramer-Rao bound

$$\sigma_{\text{TDOA}} = \frac{\sqrt{3}}{\pi B_s \sqrt{B_{\text{rcv}} T \gamma_{\text{SNR}}}} \quad (47-1)$$

where B_s is the signal bandwidth, B_{rcv} is the receiving system bandwidth, T is correlation time, and γ_{SNR} is the input correlation SNR. γ_{SNR} is derived from the sensor site pair contributing to the correlation

$$\frac{1}{\gamma_{\text{SNR}}} = \frac{1}{2} \left[\frac{1}{\text{SNR}_1} + \frac{1}{\text{SNR}_2} + \frac{1}{\text{SNR}_1 \cdot \text{SNR}_2} \right] \quad (47-2)$$

When the site pair signals are well above the noise floor, the effective γ_{SNR} is just the average of the two intercepts and the cross term $1/\text{SNR}_1 \text{SNR}_2$ can be ignored. Under low SNR conditions, all terms must be considered, and under very low SNR conditions, the cross term of Eq. 47-2 dominates. If one of the site pairs has a strong signal intercept, the correlation process can “pull out” the signal from the second site even when the signal is buried in noise (see Table 47-1).¹⁹

TABLE 47-1 SNR Input, Correlation Gain, and TDOA Measurement Error

Sensor 1 SNR (dB)	Sensor 2 SNR (dB)	Effective Input SNR to Correlator (dB)	Time-Bandwidth Ideal Correlation Gain (dB)	Correlator Peak SNR (dB)	TDOA error (ns)
20	20	20.0	37.0	57.0	16
20	12	14.3	37.0	51.3	30
20	4	6.9	37.0	43.9	71
20	0	2.9	37.0	39.9	111
20	-4	-1.0	37.0	35.9	176

System Calibration and Characterization

A core set of DF processing algorithms relies on the existence of a well-defined array manifold. The array manifold describes the antenna voltage response for an arbitrary N -element array from a set of M signal azimuth angles θ . This can be written in a matrix as

$$\mathbf{a}(\theta) = \begin{bmatrix} a_1(\theta_0) & a_1(\theta_1) & \cdots & a_1(\theta_M) \\ a_2(\theta_0) & a_2(\theta_1) & \cdots & a_2(\theta_M) \\ \vdots & \vdots & \ddots & \vdots \\ a_N(\theta_0) & a_N(\theta_1) & \cdots & a_N(\theta_M) \end{bmatrix} \quad (47-3)$$

Actual array manifolds are composed of more terms than the idealized response to azimuth described in Eq. 47-3. Array manifolds are taken at discrete steps of azimuth over many frequencies (typically at steps ~5 percent of the frequency) and may also include the dimensions of signal elevation and polarization angle. When raw manifold data is measured on a range, the result must be smoothed and interpolated to produce a manifold with uniform azimuth steps. Metadata may be included with the manifold to indicate what frequency, elevation, polarization, or other parameters are used for the manifold creation.

The accuracy of the array manifold typically dominates the system error budget and establishes DF performance. The design of the array can mitigate manifold errors through careful selection of antenna element position and number of elements. Both of these choices are almost always constrained by the platform or site physical boundaries and obstacles, and by the number of processing resources available.

For fixed-site HF antenna arrays, the electrical properties of the ground plane contribute significantly to low angle signal response, particularly the signal magnitude. For this reason, phase-only solutions should be considered. Site errors can be caused by array element interaction with adjacent antennas and nearby structures (mutual coupling, multipath, and blockage), and terrain irregularity (ground plane discontinuities). More information on site errors can be found in Gething.²⁰

For transportable or mobile systems, the arrays may need preliminary Numerical Electromagnetics Code (NEC) modeling for antenna position selection. Many times there is a difference between NEC modeling and actual field measurements (see Figure 47-3).

As expected, accurate measurements are typically difficult to obtain near resonant frequencies when the array antennas electrically couple with other antennas or structures. There is an unwritten law of DF that frequency ranges which are difficult to calibrate tend

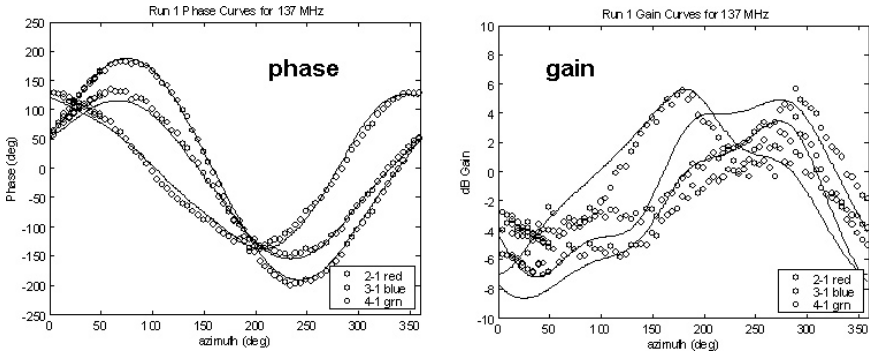


FIGURE 47-3 (a) VHF array phase and (b) gain comparison between NEC and measurement

to match the ranges where the greatest manifold accuracy is needed. Range calibration planning is required to ensure that all frequencies, azimuth and elevation angles, and signal polarizations are adequately covered.

The process of array manifold generation is performed using “collect” and “verify” functions. For the “collect,” test emissions are generated at the frequencies and polarizations of interest and the antenna array is rotated and/or tilted for the range of azimuths and elevations of interest. This may be done at a test range where the antenna array is tilted and rotated, or done on a ship or aircraft moving in a circle, arc, or other pattern at a distance from the test emitter. DF array voltages are recorded in approximately equal steps (e.g. every 1° in azimuth and at FCC allowable frequencies), smoothed, and interpolated to the final array manifold.

For the “verify,” the test emitter repeats a subset of the “collect” frequencies and polarizations. Additional frequencies are also included. The DF system receives and processes these emissions using the prepared “collect” array manifolds and computes the LOB rms error to the test emitter. Figure 47-4a illustrates LOB error magnitude versus both azimuth (y-axis degrees) and frequency (x-axis MHz). Figure 47-4b plots aggregate rms error (y-axis degrees) across the azimuth range versus frequency (x-axis MHz). The LOB accuracies expected from the system and array manifold are determined and verified in this fashion.

Built-in system features that facilitate testing of individual antenna elements, cabling loss, time delay and phase error, and overall system processing are vital to long-term AOA direction finding and TDOA geolocation effectiveness. Without such features, degradations in antenna dielectric or temperature changes in electrical properties go uncalibrated. Since DF systems often coexist with other electronic equipment, antenna element responses may change when nearby equipment and structures change. Degradations are manifested as a lack of DF productivity or increase in wild LOB estimates. Antenna designs need to include the ability to inject, couple, and radiate calibration signals without impacting voltage standing wave ratio (VSWR) and reliability.

47.3 DIRECTION FINDING SYSTEM EXAMPLES

This section examines the radio DF systems of the classic Wullenweber CDAA, the popular amateur pseudo-Doppler DF system, the historic Watson-Watt DF system, variations of the *N*-channel DF system, and Butler DF using antenna phase combining.

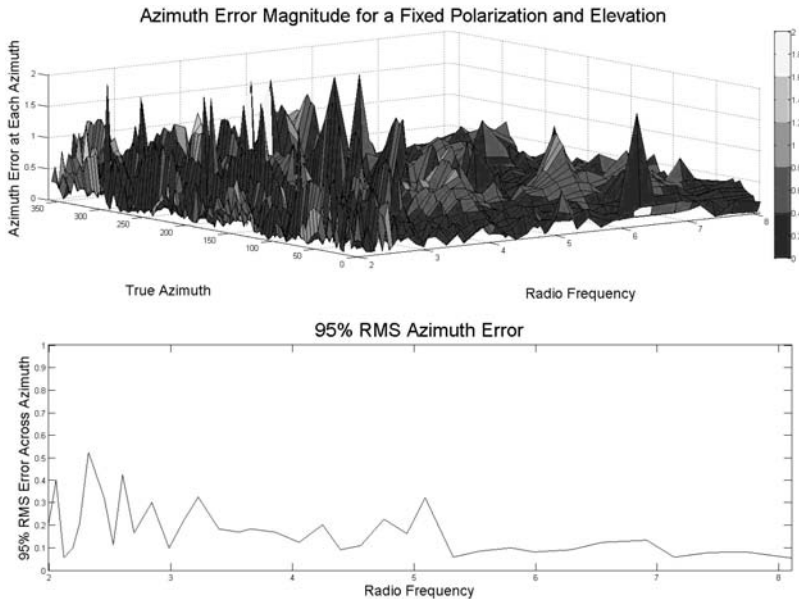


FIGURE 47-4 (a) Azimuth error vs. azimuth and frequency and (b) mean azimuth error vs. frequency

CDAA Goniometer Direction Finding System

During the 1950s the University of Illinois improved the HFDF Wullenweber CDAA for the U.S. Navy. A ring of 120 monopole antennas was cabled to a central operations building for HF analog beam combining. In the building, matched cables from the antennas were brought together at a circular stator. A spinning rotor, called a *goniometer*, capacitively coupled to 36 contiguous elements of the stator at any moment in time provided both summed and differenced outputs. Thus, the goniometer acted both as a commutator sequentially switching between antenna elements and as a beamformer.

The goniometer delay lines (see Figure 47-5) were electrical lengths equal to the free-space path of the signal to the antenna elements referenced to the center of the goniometer.

The HFDF operator could select either the sum or the difference output on a CRT display with a synchronized rotating time base so that the response pattern of the antenna appeared as a polar beam pattern centered on the direction of the emitter. When searching for an emitter, the sum mode was used, but when the emitter had been identified, the difference mode was used so that the sharp null response pattern displayed azimuth angle with maximum accuracy.

Pseudo-Doppler Direction Finding System

The pseudo-Doppler DF system is a phase-measurement system that uses commutation and stems from the research of Earp and Godfrey in 1947.²¹ This DF system is sometimes described in terms of a rotating antenna and the Doppler effect. However, the antennas do not move and there is no change of frequency due to true Doppler ($\Delta f = v f / c$). Rather, this is a clever system that measures signal phase changes using commutated (switched) sequential elements located on the circumference of a circle.

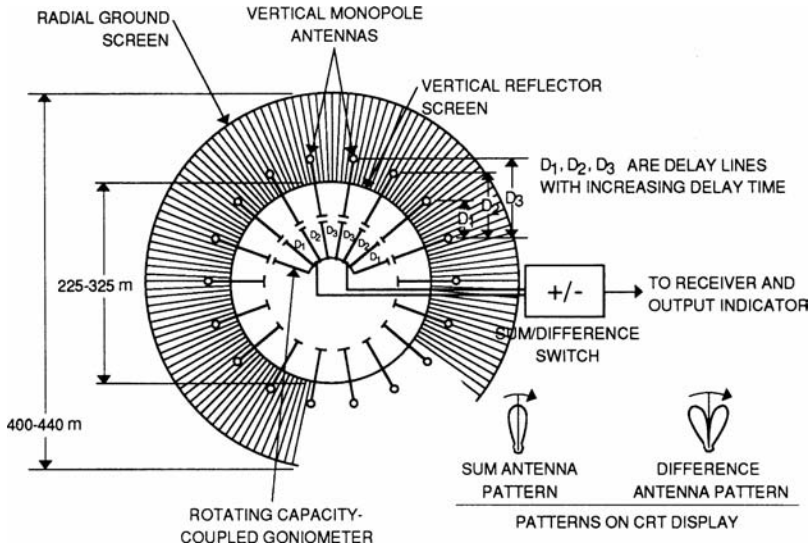


FIGURE 47-5 CDAA array goniometer beamforming

As each antenna element is sampled, the signal phase appears spatially shifted. The pseudo-Doppler DF system (see Figure 47-6) relies on a very short sampling time from each element. This is done by using a simple pin-diode set of switches that funnels each antenna element sequentially into a common RF path. The spatially induced phase shifts become manifest in the FM demodulator as slightly different audio tones. The audio is

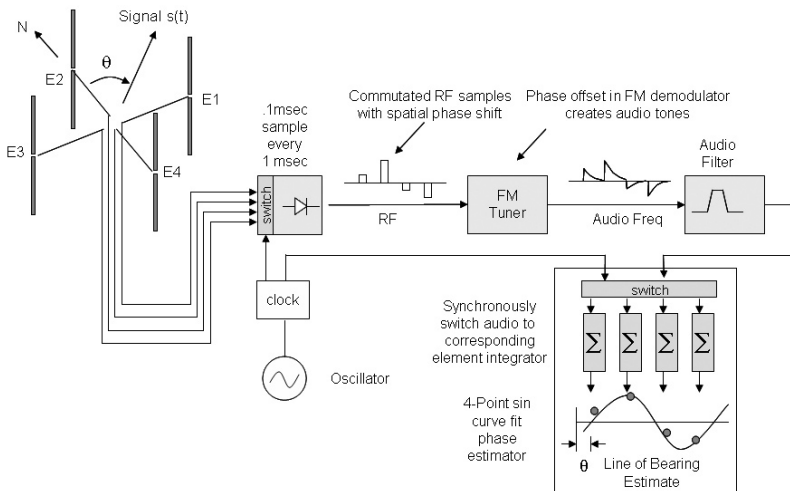


FIGURE 47-6 Pseudo-Doppler DF system using phase commutation

integrated into different summing units, commutating them synchronously at the same clocking speed that sampled the antenna elements. The integrated voltages are fitted with a sine curve whose phase is directly proportional to signal azimuth θ .

Watson-Watt Direction Finding System

The Watson-Watt DF system is a simple, but continuous, amplitude comparison scheme using two orthogonal Adcock beamforming arrays.²² Each Adcock array is made of two dipole elements that are phase-reversed to create a figure “8” beam pattern (see Figure 47-7). The optimal array spacing is about $\lambda/8$ with relatively short (0.1λ) dipole element lengths.

The Watson-Watt Adcock DF process (see Figure 47-8) uses two orthogonal Adcock arrays nicknamed the “sine” and “cosine” receiver channels, where the angle of arrival θ is estimated by taking the arctangent of the voltages from these two RF channels.²³

$$\theta = 90 - \arctan\left(\frac{V_{\sin}}{V_{\cos}}\right) \quad (47-4)$$

The Watson-Watt DF process has two fundamental problems: first, the correct quadrant of the azimuth cannot be determined uniquely, and second, the arctangent approximation deviates as the array diameter to signal wavelength (D/λ) ratio changes. That is, Eq. 47-4 is only an approximation to the true relationship between the two antenna array voltages, which is fully given as

$$\frac{V_{\sin}}{V_{\cos}} = \frac{V(E_2 - E_4)}{V(E_1 - E_3)} = \frac{\sin\left(\frac{\pi D}{\lambda} \sin(\theta)\right)}{\sin\left(\frac{\pi D}{\lambda} \cos(\theta)\right)} \quad (47-5)$$

The LOB error in using the Watson-Watt arctangent approximation is illustrated in Figure 47-9.

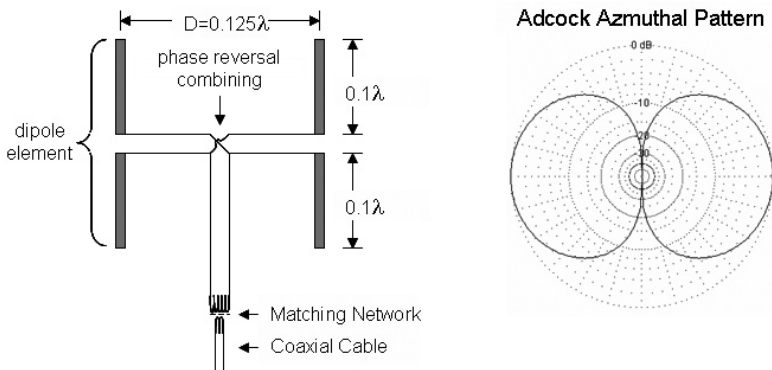


FIGURE 47-7 Adcock antenna and azimuthal beam pattern

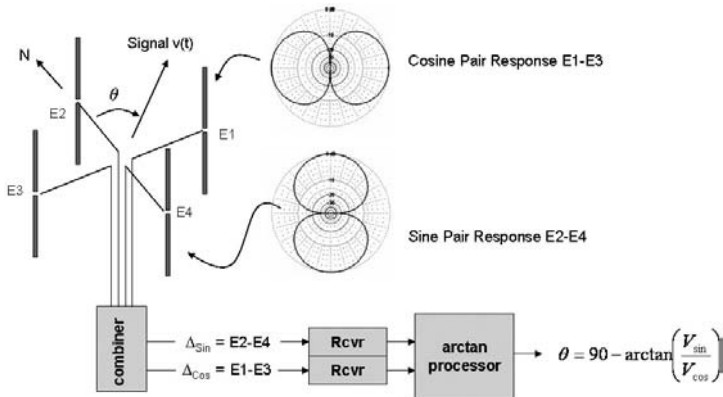


FIGURE 47-8 Watson-Watt DF system

Brueninger Direction Finding System

To correct the deficiencies of Watson-Watt DF, the Brueninger DF system makes a number of improvements:

1. A sense (reference) antenna is used to distinguish signal phase that resolves LOB quadrant ambiguities. The omni antenna may be a separate antenna or can be formed from the sum of all four Adcock elements.

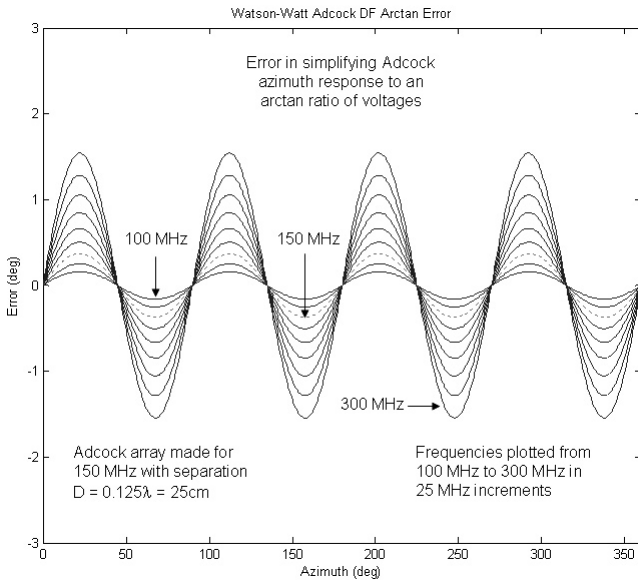


FIGURE 47-9 Watson-Watt DF error

2. More than two antenna elements may be combined to create the sine and cosine beam patterns (e.g. a symmetric set of 8 or 16 dipoles is common).
3. Explicit phase-amplitude correlation methods are used instead of taking the arctangent of voltages (e.g. Bartlett correlation DF is common).

The Brueninger DF uses three RF input channels: omni, sine, and cosine channels. Modern systems use A/D voltage conversion for digital RF processing, as shown in Figure 47-10.

In this example the DF processor implements the Bartlett correlation algorithm exploiting both the phase and amplitude of the voltages. The RF omni, sine, and cosine beam pattern complex voltages ($1 + jQ$, or “IQ” for short) are collected into an array manifold over 0° to 360° in azimuth.

To determine the direction of a signal, the voltages v_o , v_s , and v_c from the omni, sine, and cosine RF channels are measured in K snapshots, forming a voltage matrix representing all data. The voltage matrix is then used to form a time-integrated covariance matrix, \mathbf{R} :

$$\mathbf{v} = \begin{bmatrix} v_o(1) & v_o(2) & \cdots & v_o(K) \\ v_s(1) & v_s(2) & \cdots & v_s(K) \\ v_c(1) & v_c(2) & \cdots & v_c(K) \end{bmatrix} = \begin{bmatrix} \hat{v}_o \\ \hat{v}_s \\ \hat{v}_c \end{bmatrix} \quad (47-6)$$

where $v_n(k)$ indicates the k th snapshot of complex voltages from element n = omni, sine, and cosine and \hat{v}_n is the time set of voltages with time indices suppressed.

The covariance matrix (Eq. 47-7) is the product of the voltage matrix \mathbf{v} and its Hermitian transpose \mathbf{v}^H and is normalized by $1/K$ samples:

$$\mathbf{R} = \frac{1}{K} \mathbf{v} \mathbf{v}^H = \frac{1}{K} \begin{bmatrix} \hat{v}_o \hat{v}_o^T & \hat{v}_o \hat{v}_s^T & \hat{v}_o \hat{v}_c^T \\ \hat{v}_s \hat{v}_o^T & \hat{v}_s \hat{v}_s^T & \hat{v}_s \hat{v}_c^T \\ \hat{v}_c \hat{v}_o^T & \hat{v}_c \hat{v}_s^T & \hat{v}_c \hat{v}_c^T \end{bmatrix} \quad (47-7)$$

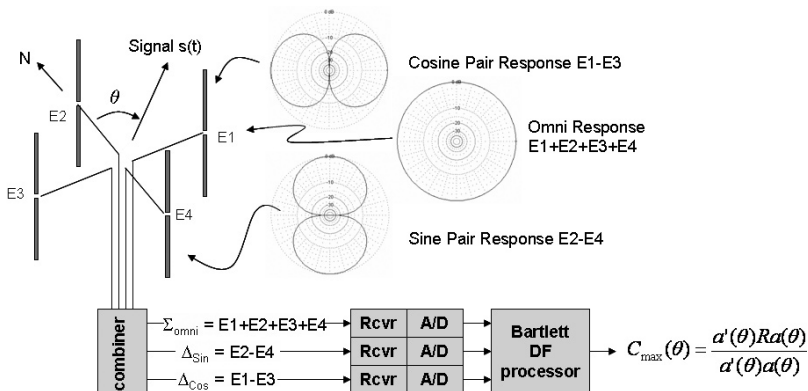


FIGURE 47-10 Brueninger omni, sine, and cosine DF system

Bartlett correlation uses the covariance matrix \mathbf{R} and the corresponding array manifold $\mathbf{a}(\theta)$ (interpolated to signal frequency) to form a correlation coefficient $C(\theta)$ for every discrete θ in the antenna manifold:

$$C(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{R} \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{a}(\theta)} \quad (47-8)$$

The coefficient $C(\theta)$ can be normalized to values 0-1 by multiplying by $1/\text{trace}(\mathbf{R})$. Once $C(\theta)$ is formed, the Bartlett DF process looks for the maximum $C(\theta)$ using standard peak finding and interpolation techniques (see Figure 47-11a). In a simulation, all signal azimuth angles can be examined to create a correlation surface (see Figure 47-11b). With a VHF four-element Adcock array 40 cm in diameter, the bearing error at 150 MHz is about 2° rms.

Commuted Direction Finding Systems

As described earlier in this section, the CDAA and pseudo-Doppler DF systems use antenna element commutation. The Brueninger DF system and other arbitrary DF arrays can also be commuted. One antenna element (usually the “omni”) is selected as the reference element to which all other antenna elements are sequentially compared. Although the array manifold remains the same, the covariance matrix becomes sparse. For example, if antenna element #1 is the reference element, the covariance matrix for an N -element array becomes

$$\mathbf{R} = \frac{1}{K} \begin{bmatrix} \hat{\mathbf{v}}_1 \hat{\mathbf{v}}_1^T & \hat{\mathbf{v}}_1 \hat{\mathbf{v}}_2^T & \cdots & \hat{\mathbf{v}}_1 \hat{\mathbf{v}}_N^T \\ \hat{\mathbf{v}}_2 \hat{\mathbf{v}}_1^T & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{v}}_N \hat{\mathbf{v}}_1^T & 0 & \cdots & 0 \end{bmatrix} \quad (47-9)$$

The Bartlett Eq. 47-8 is modified to the commuted form

$$C(\theta) = \frac{\mathbf{R}^H \mathbf{a}(\theta)}{[\mathbf{a}(\theta)^H \mathbf{a}(\theta)]^{\frac{1}{2}}} \quad (47-10)$$

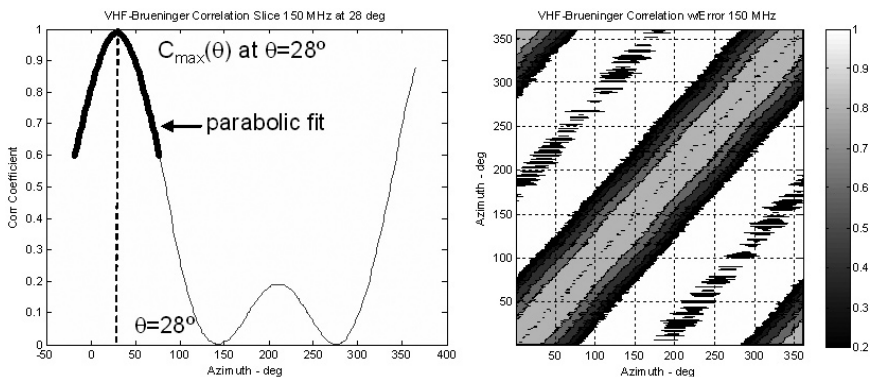


FIGURE 47-11 (a) Bartlett correlation with peak finding and (b) Bartlett correlation surface

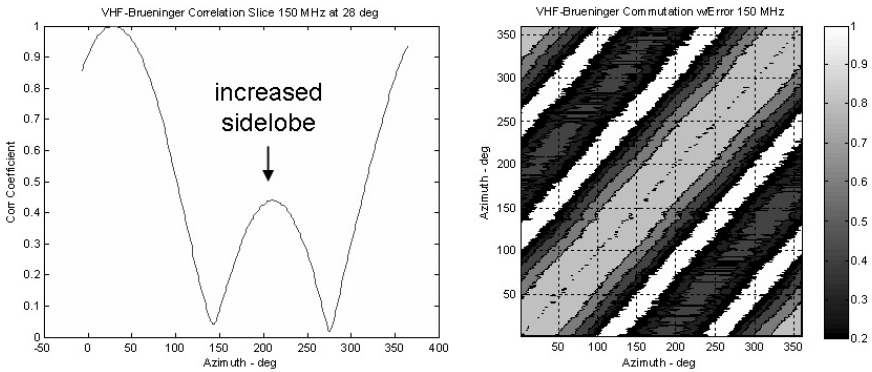


FIGURE 47-12 VHF Brueninger commutated DF (a) correlation with signal at 28° and (b) correlation surface

In spite of what appears to be a paucity of data, DF commutation using the Bartlett algorithm performs nearly as well as the full N -channel solution provided that the total number of voltage samples is the same. However, an N -element commutated array takes N times longer to collect the same number of RF voltage samples. For example, the Brueninger omni, sine, cosine DF system performing three-channel correlation (see Figure 47-11) can be compared to two different commutation variations: a Brueninger commutated system (see Figure 47-12) and a four-element array that is commutated directly without Adcock element combining (see Figure 47-13). As in the previous section, the array example contains four VHF dipoles on a 40-cm circle operating at 150 MHz. Notice that for Brueninger commutation, there is an increase in spatial sidelobes that is not present in four-element sampling.

Why commute? Because it reduces the amount of DF hardware by a factor of $2:N$, which for an 8- or 16-element array can be significant. Regardless of array size, the commutating system only requires two channels of filters, tuners, and A/D converters. Similar efficiencies are gained in digital signal processing. However, the commutation system pays a penalty in time to collect the same number of data snapshots and may totally corrupt the covariance matrix if either the signal transmission is shorter than the commutation cycle time or two or more signals occur at the same frequency during the commutation cycle time.

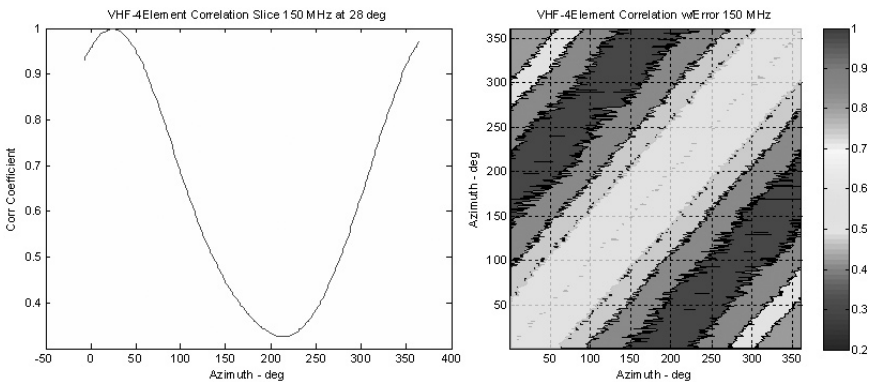


FIGURE 47-13 VHF four-element commutated DF (a) correlation with signal at 28° and (b) correlation surface

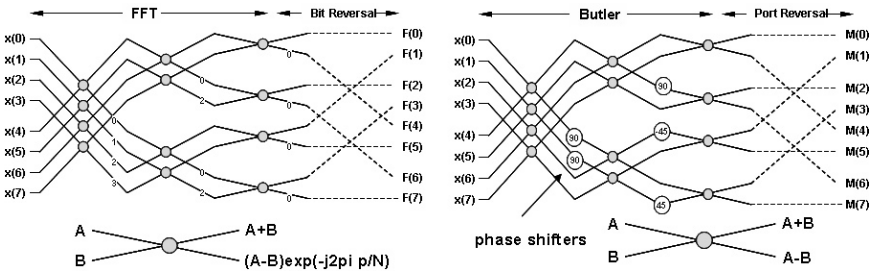


FIGURE 47-14 (a) Fast Fourier transform and (b) Butler Matrix antenna combining

Butler Direction Finding System

When discussing antenna element combining, people often confuse the Brueninger and Butler combiners. The Brueninger, even in its most complex form, only phase reverses antenna elements by 180° before combining, and the result is always three outputs of an omni, sine, and cosine phased beam pattern. Butler combining, most often implemented as a “Butler Matrix” (see Figure 47-14b), is more akin to fast Fourier transform (FFT) and uses combinations of 45° , 90° , and 180° phase shift components.

The FFT example above shows eight inputs and the corresponding eight outputs after “bit reversal” that reorders the output. If the input is time-ordered data, the output is frequency-ordered data with $F(0)$ being the DC component. Butler antenna combining follows the same “butterfly diagram.” Each node creates a sum and difference ($A + B$) and ($A - B$) without exponential phase weighting. The phase weighting comes at discrete points in the diagram, requiring 45° and 90° phase shifters. If the input is RF energy from eight equally spaced antennas around a circle, the output is spatial phase shifted modes. Mode 0 has no spatial phase shifting with respect to signal azimuth and is exactly equivalent to an omnidirectional beam combiner. Mode 1 and Mode 7 (referred to as Mode -1) have spatial phase shifting proportional to signal azimuth. Similarly Mode 2 and 6 (Mode -2) have spatial phase shifting proportional to twice the signal azimuth. Mode 3 has spatial phase shifting proportional to three times the signal azimuth.

Figure 47-15a shows a Monte Carlo simulation of Mode 0 and Mode 1 while Figure 47-15b shows the corresponding Butler DF response using the same notional VHF four-element array used earlier in Section 47.3.

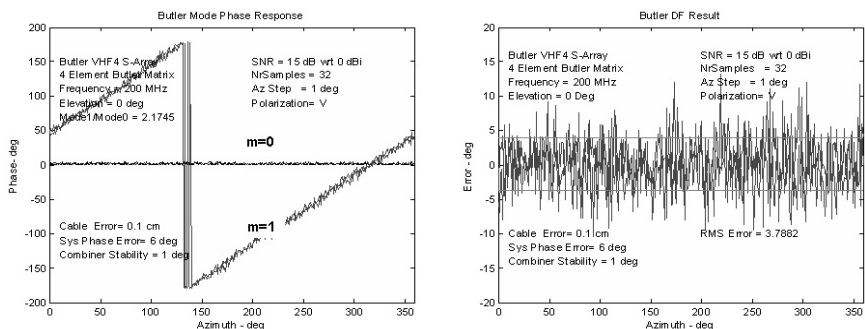


FIGURE 47-15 (a) Butler Mode 0 and 1 and (b) resulting DF output

The Butler system uses Mode 0 for signal acquisition and reference to which other modes are phase compared. Mode 1 is typically used for coarse DF with the DF algorithm switching to Mode 2 and even Mode 3 for higher resolution. The DF cascade is necessary because Mode 2 and Mode 3 by themselves are double and triple valued over the full azimuth range and need guidance to signal direction from the previous stage.

47.4 DIRECTION FINDING ALGORITHMS

Modern radio DF systems are dependent upon the direction finding algorithm. Section 47.3 introduced the Bartlett DF algorithm, but it is just one of a host of DF solutions. In this section, we consider several classes of DF algorithms, which include correlation-based, eigen-structured, root-finding, and maximum-likelihood algorithms. We then introduce the Cramer-Rao bound (CRB) that specifies the lower bound for DF error due to additive white Gaussian noise. We then compare various DF algorithms against the CRB. Lastly, we consider the effects of correlated cochannel interference on DF performance.

Basic Correlation Algorithms

We have already introduced the Watson-Watt voltage comparison algorithm and Bartlett correlation for Brueninger and commutated arrays. Here we present a comparison of generic DF algorithms that can be applied to any arbitrary array geometry. Two matrices define the environment that the DF algorithms operate in. The first is \mathbf{a} , the N -element by M -azimuth array manifold that specifies the complex response of each of the N elements over the M azimuths in the field of regard. The second is the sampled signal voltage \mathbf{v} , which is an $N \times K$ matrix, where K is the number of signal samples:

$$\mathbf{a}(\theta) = \begin{bmatrix} a_1(\theta_0) & a_1(\theta_1) & \cdots & a_1(\theta_M) \\ a_2(\theta_0) & a_2(\theta_1) & \cdots & a_2(\theta_M) \\ \vdots & \vdots & \ddots & \vdots \\ a_N(\theta_0) & a_N(\theta_1) & \cdots & a_N(\theta_M) \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1(0) & v_1(1) & \cdots & v_1(K) \\ v_2(0) & v_2(1) & \cdots & v_2(K) \\ \vdots & \vdots & \ddots & \vdots \\ v_N(0) & v_N(1) & \cdots & v_N(K) \end{bmatrix} \quad (47-11)$$

The covariance matrix remains defined as $\mathbf{R} = \frac{1}{K} \mathbf{v} \mathbf{v}^H$. Using these definitions, various correlation DF algorithms have been created.^{24,25}

$$\text{Bartlett (Normalized):} \quad C(\theta) = \left(\frac{1}{\text{tr}(\mathbf{R})} \right) \frac{\mathbf{a}(\theta)^H \mathbf{R} \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{a}(\theta)} \quad (47-12)$$

$$\text{Capon (Classic):} \quad C(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{R}^{-1} \mathbf{a}(\theta)} \quad (47-13)$$

$$\text{Capon (Smooth):} \quad C(\theta) = 1 - \frac{\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{a}(\theta)} \quad (47-14)$$

$$\text{Maximum Entropy:} \quad C_k(\theta) = \frac{\mathbf{a}(\theta)^H \mathbf{a}(\theta)}{\mathbf{a}(\theta)^H \mathbf{C}_k^{-1} \mathbf{a}(\theta)} \quad (47-15)$$

The performance of these algorithms will be given later in the chapter, in the section “DF Algorithm Comparison.” For the Bartlett algorithm, we included the factor $1/\text{trace}(\mathbf{R})$ to ensure $C(\theta)$ varies between zero and one. The Capon algorithm is written in both its classic form and in a smooth (inverted) form that ensures $C(\theta)$ varies between zero and one and allows easier peak interpolation when using discrete array manifold θ steps. For the Maximum Entropy algorithm, \mathbf{c}_k is the k th column vector of the inverse covariance matrix (\mathbf{R}^{-1}) .

Eigen Structured Algorithms

Many DF algorithms depend upon decomposing the covariance matrix into subspaces that comprise the signal and noise. One of the first spectral estimation methods using eigenvectors was developed in 1973 by V. F. Pisarenko,²⁶ whose eigen method originally focused on harmonic decomposition. The Min-Norm eigen approach was developed in 1983 by Kumaresan and Tufts.²⁷ However, the most popular and versatile algorithm is Schmidt’s 1986 MUSIC algorithm.¹⁵ The algorithms are

Pisarenko:
$$C_k(\theta) = \frac{1}{|\mathbf{a}(\theta)^H \bar{\mathbf{e}}_k|^2} \quad (47-16)$$

Min-Norm:
$$C_k(\theta) = \frac{(\mathbf{u}_k^T \mathbf{E}_N \mathbf{E}_N^H \mathbf{u}_k)^2}{|\mathbf{a}(\theta)^H \mathbf{E}_N \mathbf{E}_N^H \mathbf{u}_k|^2} \quad (47-17)$$

MUSIC (classic):
$$C(\theta) = \frac{1}{\mathbf{a}(\theta)^H \{\mathbf{E}_N \mathbf{E}_N^H\} \mathbf{a}(\theta)} \quad (47-18)$$

MUSIC (smooth):
$$C(\theta) = 1 - \frac{\mathbf{a}^H(\theta) \{\mathbf{E}_N \mathbf{E}_N^H\} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{a}(\theta)} = \frac{\mathbf{a}^H(\theta) \{\mathbf{E}_S \mathbf{E}_S^H\} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{a}(\theta)} \quad (47-19)$$

The performance of these algorithms will also be given later in the chapter, in the section “DF Algorithm Comparison.” For the Pisarenko algorithm, $\bar{\mathbf{e}}_k$ is one or more eigenvectors associated with the smallest (noise) eigenvalues. The span of the eigenvectors is of dimension N , the number of elements in the array. We normally expect only one signal present for DF. Hence the signal eigenvector \mathbf{E}_S is of dimension $N \times 1$, and \mathbf{E}_N , the matrix of noise eigenvectors, is of dimension $N \times (N - 1)$. If D co-channel signals are present in the voltage samples that make up the covariance matrix, \mathbf{E}_S is an $N \times D$ subspace and \mathbf{E}_N is an $N \times (N - D)$ subspace.

For the Min-Norm algorithm, \mathbf{u}_k is the k th column of an $N \times N$ identity matrix. The MUSIC algorithm is presented in its classical form in Eq. 47-18, which shows distinctive sharp spectral peaks, and is presented in a smoothed form in Eq. 47-19, which has smoother peaks and is easier to interpolate when using a discrete array manifold. Further, MUSIC is written using either the noise or signal eigenvectors \mathbf{E}_N or \mathbf{E}_S : choose the form that has the smaller subspace for the lowest computational load.

Root DF Algorithms

Root DF algorithms are specialty DF algorithms applicable to only uniformly spaced linear arrays with N elements, where the array manifold can be written using an index to specify the array element of interest. The array manifold is then written as

$$\mathbf{a}(\theta, n) = \exp\left(-jn \frac{2\pi d}{\lambda} \cos(\theta)\right) \quad (47-20)$$

where n goes from $-(N-1)/2$ to $(N-1)/2$ (odd-numbered element arrays), λ is the signal wavelength, and d is the spacing between elements. For a standard linear array, just at the Nyquist limit, $d = \lambda/2$. Then Eq. 47-20 reduces to

$$\mathbf{a}(\theta, n) = \exp(-jn\pi \cos(\theta)) \quad (47-21)$$

The array manifold can further be transformed into a vector of polynomial coefficients z by using the representation

$$z = \exp(j\pi \cos(\theta)) \quad \text{giving} \quad \mathbf{a}_z(z) = [1 \quad z \quad \dots \quad z^{N-1}]^T \quad (47-22)$$

The most common DF approach is the Root MUSIC algorithm using noise subspace eigenvectors. The correlation coefficient is written in terms of the eigenvector noise subspace and the antenna manifold as

$$C(z) = \mathbf{a}^T \left(\frac{1}{z} \right) \{ \mathbf{E}_N \mathbf{E}_N^H \} \mathbf{a}(z) \quad (47-23)$$

The Root MUSIC algorithm computes the roots of $C(z)$ and chooses the D roots, \hat{z}_i , that lie inside a unit circle and are closest to the unit circle. The estimates of signal azimuth $\hat{\theta}_i$ are found by the inverse relation between θ and z for the $i = 1 \dots D$ roots, giving

$$\hat{\theta}_i = \cos^{-1} \left(\frac{\arg(\hat{z}_i)}{\pi} \right) \quad (47-24)$$

From Eq. 47-24 it can be seen that Root MUSIC is not affected by the magnitude of \hat{z}_i . Thus we expect Root MUSIC to provide a better estimate of signal direction than the classical MUSIC algorithm. The performance of these algorithms is given later in the chapter, in the section “DF Algorithm Comparison.”

Since the Min-Norm algorithm also uses eigenvectors of the noise subspace, we expect that there should be a Root Min-Norm algorithm that is a better estimator than Eq. 47-17. While the development of both root MUSIC and Root Min-Norm is developed using a uniformly spaced linear array, these algorithms are applicable to circular arrays as well. The starting point for Root Min-Norm begins by defining a subspace vector \mathbf{d} , of dimension $N \times 1$, that has the property of being orthogonal to the signal subspace. That is, given the array manifold $\mathbf{a}(\theta_i)$ at discrete θ_i for $i = 1 \dots D$ signal directions, \mathbf{d} has the property

$$\mathbf{a}(\theta_i)^H \mathbf{d} = 0 \quad (47-25)$$

The only set of vectors orthogonal to the signal vectors is the set of noise eigenvectors with arbitrary magnitude. Thus the vector \mathbf{d} is found as a weighted noise subspace eigenvector

$$\mathbf{d} = w^{\frac{1}{2}} \mathbf{E}_N = \begin{pmatrix} \mathbf{c}^* \\ \mathbf{c}^H \mathbf{c} \end{pmatrix} \mathbf{E}_N \quad (47-26)$$

where \mathbf{c} is created from the first elements of each of the noise eigenvectors. As with Root MUSIC, Root Min-Norm uses the transform $z_i = \exp(j\pi \cos(\theta_i))$ to represent the $i = 1 \dots D$ signal directions, allowing \mathbf{d} to be written in polynomial form

$$D(z) = 1 + d_1 z^{-1} + \dots + d_N z^{-N} \quad (47-27)$$

The polynomial roots of $D(z)$ provide the estimate of \hat{z}_i , which are selected within and close to the unit circle. The inverse trigonometric transform of Eq. 47-24 converts \hat{z}_i to the Root Min-Norm estimate of signal arrival $\hat{\theta}_i$ for the $i = 1 \dots D$ signals.

Maximum Likelihood DF Algorithms

The Maximum Likelihood (ML) estimator for the direction $\hat{\theta}_i$, for $i = 1 \dots D$ signals, uses a signal subspace that is as close as possible to the data subspace. This is formulated by taking the maximum of

$$\theta_{\text{ml}} = \arg \max_{\theta} \{tr[\mathbf{P}_{\theta} \mathbf{R}]\} \quad (47-28)$$

where \mathbf{R} is the covariance matrix of the antenna array voltage data and \mathbf{P}_{θ} is the projection vector of the $i = 1 \dots D$ signals. \mathbf{P}_{θ} is formed from the discrete $N \times D$ steering vectors $\mathbf{a}(\theta_i)$ as

$$\mathbf{P}_{\theta} = \mathbf{a}(\theta_i) (\mathbf{a}(\theta_i)^H \mathbf{a}(\theta_i))^{-1} \mathbf{a}(\theta_i)^H \quad (47-29)$$

To search over the actual N -elements by the M -azimuth array manifold $\mathbf{a}(\theta)$, the computational complexity of the ML algorithm is $O(M^D)$. Fortunately there is a way to reduce computational complexity to find \mathbf{P}_{θ} . The alternating projection maximization (APM) algorithm, invented by Ziskind and Wax,²⁸ replaces the multidimensional maximization problem with a sequence of one-dimensional maximization problems, reducing the computational complexity to $O(MD)$.

The first step of the APM algorithm is an initialization step where θ_i^0 is estimated for $i = 1 \dots D$ signals. θ_1^0 is estimated by finding the angle θ that maximizes Eq. 47-28 assuming only one signal is present. θ_2^0 is found by assuming the first signal is at θ_1^0 and maximizing Eq. 47-28 for the second signal. The initialization is complete after repeating for D signals.

Next, a projection of the manifold is created so that all of the manifold vectors are orthogonal to all of the estimated steering vectors except the steering vector that is being updated. Notationally, $\mathbf{a}(\hat{\theta}_i)$ represents the set of steering vectors for all of the estimated signals except for signal i . APM creates a projected manifold \mathbf{b} such that the vector at each θ , $\mathbf{b}(\theta, \hat{\theta}_i)$ is orthogonal to all of the vectors in $\mathbf{a}(\hat{\theta}_i)$. That is

$$\mathbf{b}(\theta, \hat{\theta}_i) \mathbf{a}(\hat{\theta}_i) = \mathbf{0} \quad (47-30)$$

APM then determines θ_i^{k+1} for iteration $k + 1$ and signal i as

$$\theta_i^{k+1} = \arg \max_{\theta_i} \mathbf{b}(\theta, \hat{\theta}_i^k)^H \mathbf{R} \mathbf{b}(\theta, \hat{\theta}_i^k) \quad (47-31)$$

The iterations are repeated until a criterion δ is reached such that

$$|\theta_i^{k+1} - \theta_i^k| < \delta \quad (47-32)$$

Power Method Algorithms for Eigen Structures

In this section we investigate ways of computing eigenvalues and eigenvectors efficiently. Many computer languages provide a Linear Algebra extension from the LAPACK library to compute eigenvectors and eigenvalues quickly.²⁹ For an N -element array, this requires $O(N^3)$ operations. For many applications, the LAPACK library meets all computational needs. Although efficient, the standard LAPACK library has a highly complex logical structure that cannot be implemented efficiently in pipeline hardware such as Field Programmable Gate Arrays (FPGAs).

The Jacobi algorithm for computing eigenvalues and eigenvectors is not optimal in terms of multiply, divide, and add operations, but it can be implemented using Coordinate Rotation Digital Computer (CORDIC) processes³⁰ that are suitable for pipeline hardware. Ahmedsaid et al³¹ implemented the Jacobi algorithm using CORDIC blocks for an 8×8 covariance matrix. Computing the eigenvectors and values to 14-bit resolution required 16,681 slices and approximately 18,000 clock cycles.

Another pipeline approach is the “power method” by Stewart,³² which shows there is an extremely simple way of estimating the dominant eigenvector, \mathbf{E}_1 , from a covariance matrix \mathbf{R} :

$$\mathbf{E}_1 = \mathbf{R}^k \mathbf{R}(m) = (\mathbf{L}\mathbf{L}^H)^k \mathbf{L}\mathbf{L}(m)^H \quad (47-33)$$

where k is an integer > 0 , $\mathbf{R} = \mathbf{L}\mathbf{L}^H$, $\mathbf{R}(m)$ is a column vector from \mathbf{R} , and $\mathbf{L}(m)$ is a row vector from \mathbf{L} . In the case of a single signal with positive SNR, \mathbf{E}_1 (an orthonormal vector) can be used in place of \mathbf{E}_s in the MUSIC DF of Eq. 47-19 and is quickly solved.

With a modest extension, the power method is able to determine a set of orthonormal vectors and bases to replace the eigenvectors and values of multiple signals. Once \mathbf{E}_1 has been found, the \mathbf{L} matrix is updated as

$$\mathbf{u} = \mathbf{E}_n^H \mathbf{L}_n \quad (47-34)$$

$$\mathbf{L}_{n+1} = \mathbf{L}_n - \mathbf{E}_n \mathbf{u} \quad (47-35)$$

where \mathbf{E}_n is normalized such that $\mathbf{E}_n^H \mathbf{E}_n = 1$. The corresponding basis value λ_n is then estimated as

$$\lambda_n = \mathbf{u}^H \mathbf{u} \quad (47-36)$$

This procedure is repeated to estimate \mathbf{E}_n and λ_n from $n = 1$ to N array elements. In all, the power method requires $(2k + 4)N^3 - (2k + 1)N^2 - 2N$ multiply and accumulate (MAC) operations. Up to $N = 8$ and $k = 2$ MAC operations can be supported using modern FPGAs.

Cramer-Rao DF Error Limit

Before evaluating each of the algorithms given in the preceding section, we would like a baseline that gives the limit of DF performance for any given array and incoming co-channel set

of signals. This can be obtained using the Cramer-Rao bound (CRB) where additive white Gaussian noise (AWGN) corrupts the received signal voltages without bias. A thorough treatment is given by Van Trees.³³

We start with a model of the received signal \mathbf{V} as an $M \times K$ matrix of voltages sampled at each of M antenna array elements over K snapshots. (We use M vice N to avoid confusion with the noise component \mathbf{N} .) \mathbf{V} is the result of D co-channel signals \mathbf{S} received through the array manifold $\mathbf{a}(\theta)$ plus direction independent noise \mathbf{N} . \mathbf{S} is a $D \times K$ matrix, while \mathbf{N} is an $M \times K$ diagonal matrix of AWGN noise samples with variance σ_w^2 .

$$\mathbf{V} = \mathbf{a}(\theta)\mathbf{S} + \mathbf{N} \quad (47-37)$$

Using this separation of signal and noise, the covariance matrix \mathbf{R}_V of measured voltages can be written in terms of signal and noise as well:

$$\mathbf{R}_V = \mathbf{R}_S + \mathbf{R}_N = \mathbf{a}(\theta)\mathbf{S}_f\mathbf{a}(\theta)^H + \sigma_w^2\mathbf{I} \quad (47-38)$$

where \mathbf{R}_S is the signal covariance matrix (equal to $\mathbf{a}(\theta)\mathbf{S}_f\mathbf{a}(\theta)^H$), \mathbf{R}_N is the noise covariance matrix (equal to $\sigma_w^2\mathbf{I}$), and \mathbf{S}_f is the $D \times D$ signal matrix (equal to $\mathbf{S}\mathbf{S}^H$). The resulting Cramer-Rao bound of variance becomes

$$\mathbf{C}_{CR}(\theta) = \frac{\sigma_w^2}{2K} \left\{ \text{Re} \left[\left[\frac{\partial \mathbf{a}(\theta)^H}{\partial \theta} \mathbf{P}_\theta^\perp \frac{\partial \mathbf{a}(\theta)}{\partial \theta} \right] \circ \left[\mathbf{S}_f \mathbf{a}(\theta)^H \mathbf{R}^{-1} \mathbf{a}(\theta) \mathbf{S}_f^T \right] \right] \right\}^{-1} \quad (47-39)$$

where \circ is the Hadamard product of two matrices using element by element multiplication and \mathbf{P}_θ^\perp is the projection matrix of the array manifold onto the noise subspace and is defined as

$$\mathbf{P}_\theta^\perp = \mathbf{I} - \mathbf{a}(\theta)(\mathbf{a}(\theta)^H \mathbf{a}(\theta))^{-1} \mathbf{a}(\theta)^H \quad (47-40)$$

Clearly the CRB decreases with increasing data samples, K , and decreasing noise variance. At high SNR the CRB is proportional to σ_w^2 , but at low SNR, the CRB is proportional to σ_w^4 . The transition from low to high SNR is determined by the signal correlation, spatial separation of the signals, and the beamwidth of the array. As the signals become more correlated, \mathbf{R}_S becomes less diagonal, increasing the CRB. As the separation of the signal steering vectors $\mathbf{a}(\theta)^H \mathbf{a}(\theta)$ becomes less diagonal, CRB likewise increases.

DF Algorithm Comparison

Monte Carlo simulations were run to compare DF performance in which two uncorrelated, equal power signals impinge on a standard linear array with elements spaced $\lambda/2$ apart. The signals were positioned nearly broadside to the array at azimuths of 87.5° and 92.5° . The array SNR (ASNR) was varied from -20 dB to $+30$ dB and resulting DF rms errors were determined from 200 trials per algorithm. Each trial uses 100 data (voltage) snapshots with additive white Gaussian noise.

If only one signal is present in the array manifold (per frequency channel, etc.), all computational methods produce the same directional estimate, with Bartlett correlation the simplest to implement. The Bartlett correlation algorithm can be used on many different array geometries, can handle Brueninger element combining, and can be used in comutating systems. But the Bartlett DF algorithm cannot resolve closely spaced co-channel

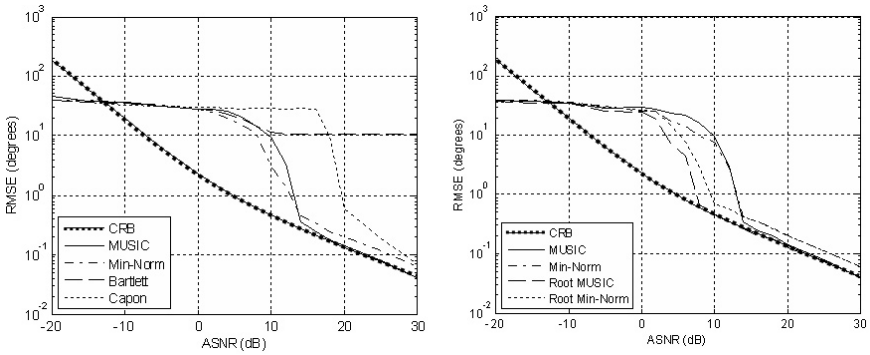


FIGURE 47-16 (a) Basic DF and Eigen algorithm DF performance and (b) Eigen and Root algorithm DF performance

signals. The Capon method provides much better signal resolution than Bartlett, but eigen structure methods provide even more capability. The Min-Norm algorithm provides slightly better performance than MUSIC at low ASNR, but with much more complexity (see Figure 47-16a). For restricted array geometry, Root MUSIC and Root Min-Norm give better DF performance than their general array counterparts MUSIC and Min-Norm. Root MUSIC is slightly better than Root Min-Norm in accuracy and is considerably more efficient to compute (see Figure 47-16b).

The power method creating orthogonal basis vectors and values can be used in the MUSIC, Root MUSIC, Min-Norm, and Root Min-Norm algorithms just as the eigenvectors and values are used. In a DF simulation consistent with the other comparisons in this section, the eigen decomposition required double floating point values (64 bits) while the power method used only 18-bit integer multipliers and 48-bit integer accumulators. The power method thus provides significant computational load savings without loss of accuracy (see Figure 47-17).

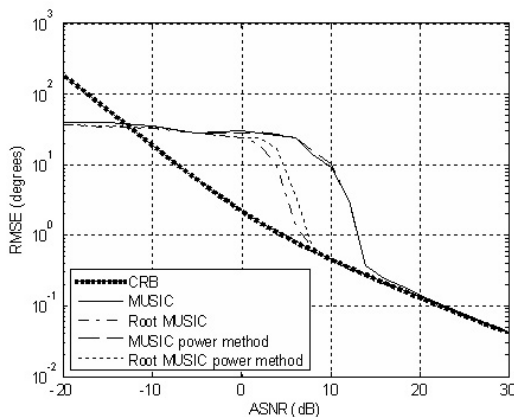


FIGURE 47-17 DF performance of Eigen versus power method implementations

Estimating the Number of Co-Channel Signals

For an N -element array, eigen decomposition of the covariance matrix gives N eigenvalues λ and corresponding eigenvectors \mathbf{E} . The eigen subspace is divided into \mathbf{E}_S , a subspace of $1 \dots D$ signal eigenvectors, and \mathbf{E}_N , a subspace of $(N - D)$ noise vectors. These eigenvectors and their eigenvalues are related to the original covariance matrix as

$$\mathbf{R} = \mathbf{E}_S \lambda_S \mathbf{E}_S^H + \mathbf{E}_N \lambda_N \mathbf{E}_N^H \quad (47-41)$$

The problem is to estimate the number of signals D when the noise eigenvalues λ_N can be represented by a span of $(N - D)$ noise terms containing AGWN,

$$\lambda_N = \text{diag}[\sigma_w^2 \dots \sigma_w^2] \quad (47-42)$$

An ad hoc ratio test can be established to compare the smallest eigenvalue with the $(N - 1)$ remaining eigenvalues in a threshold test. An optimal approach uses the likelihood ratio that challenges the hypothesis that the $(N - D)$ smallest eigenvalues are equal against the hypothesis that the $(N - D - 1)$ smallest eigenvalues are equal. Van Trees³⁴ shows that for a covariance matrix \mathbf{R} , created from K snapshots and decomposed into N eigenvalues, the likelihood of D signals, $L(D)$, can be written as

$$L(D) = K(N - D) \ln \left\{ \frac{\frac{1}{N - D} \sum_{i=D+1}^N \lambda_i}{\left(\prod_{i=D+1}^N \lambda_i \right)^{\frac{1}{N - D}}} \right\} \quad (47-43)$$

The term in the brackets is the ratio of the arithmetic mean of the $(N - D)$ smallest eigenvalues to the geometric mean of $(N - D - 1)$ eigenvalues. If the $(N - D)$ smallest eigenvalues are equal (i.e., have the same noise variance σ_w^2) then the arithmetic and geometric means are equal and $L(D) = 0$.

Two common tests use this likelihood ratio to determine the number of signals present: the Akaike information-theoretic criterion (AIC)³⁵ and the minimum description length (MDL) of Wax and Kailath.³⁶ These tests add a penalty function based on the degrees of freedom of the likelihood function. Akaike introduces AIC as

$$D_{\text{AIC}} = \arg \min_D \{L(D) + [D(2N - D)]\} \quad (47-44)$$

while the MDL test uses

$$D_{\text{MDL}} = \arg \min_D \left\{ L(D) + \frac{1}{2} [D(2N - D) + 1] \ln K \right\} \quad (47-45)$$

As with any eigen structure method, the power method of determining orthonormal vectors and values may be substituted. For the antenna array and simulation parameters defined in the preceding section, the AIC test (using eigen structures or orthonormal vectors) has a better probability of signal detection at low ASNR than the MDL test. Both AIC and MDL may underestimate the number of signals present. Above 8 dB, all methods successfully estimate the presence of two signals (see Figure 47-18).

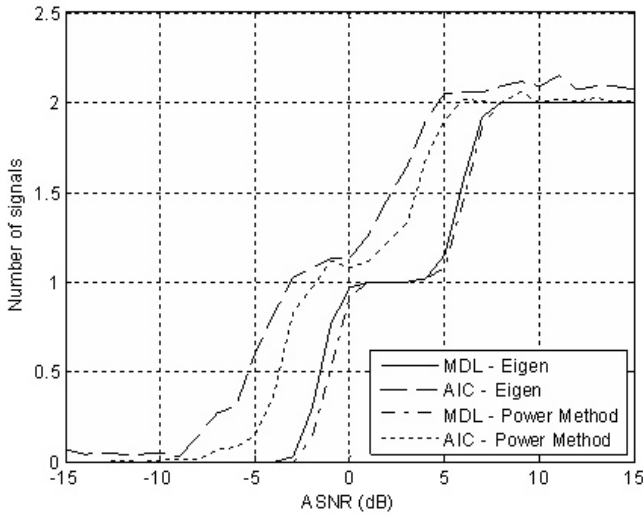


FIGURE 47-18 Estimates of the number of signals present vs. array signal-to-noise ratio (ASNR)

DF when Co-Channel Signals Are Correlated

All of the DF algorithms discussed so far have some ability to separate co-channel signals. But none can separate perfectly correlated or coherent signals. They cannot separate the signals because the signal covariance matrix, S_f , as defined in Eq. 47-38, has a rank of one for coherent signals.

Unfortunately, multiple signal correlation generally occurs in a multipath situation. That is, a single signal travels on multiple paths from the transmitter to the receiving antenna array. If the signal arrives on two paths that are equal length, then the antenna array would see two signals that are 100 percent correlated. As the path length difference increases, the correlation ratio drops. An estimate of the correlation, ρ , is given

$$\rho = 1 - \frac{B \cdot \Delta path}{c} \quad (47-46)$$

where B is signal bandwidth, $\Delta path$ is the path length difference, and c is the velocity of light. For example, if the signal bandwidth is 25 kHz and the path length difference is 600 meters, then $\rho = 0.95$.

A Monte Carlo simulation similar to that used previously in Section 47.4 was used to evaluate MUSIC and Root MUSIC using two signals with 95 percent correlation (see Figure 47-19). At low ASNR, both MUSIC and Root MUSIC are erroneous, randomly determining the DF. Even a moderately strong ASNR of 10–12 dB still produces random DF. Only when the SNR is greater than 15 dB does Root MUSIC begin to determine the correct direction of both signals, while the classic MUSIC algorithm needs ASNR > 25 dB to produce useful DF. Both Root MUSIC and classic MUSIC can resolve highly correlated, closely spaced signals given sufficient ASNR.

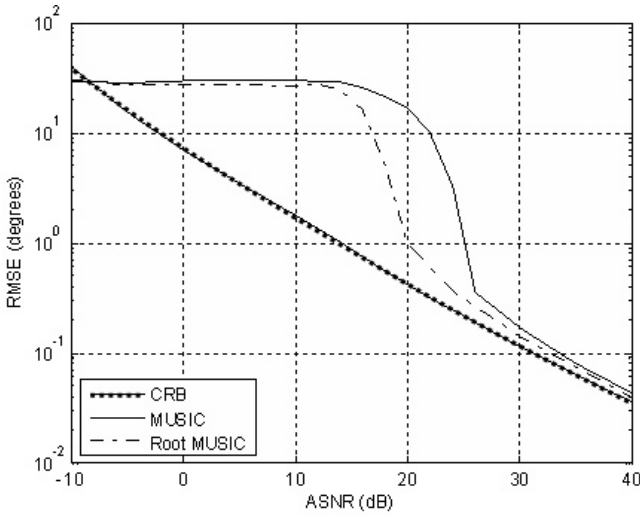


FIGURE 47-19 DF performance for two closely spaced signals that are 95 percent correlated

47.5 GEOLOCATION ALGORITHMS

In this section, we present three geolocation algorithms. The Stansfield algorithm is the oldest and has been used successfully in both worldwide HFDF systems and line-of-sight angle-of-arrival systems as well. The Wangness algorithm uses LOBs drawn on a spherical earth and likewise can be used globally or locally for geolocation solutions. A simple 2D TDOA geolocation algorithm is also presented. These three approaches are compared using confidence ellipse containment.

Stansfield AOA Geolocation Algorithm

The Stansfield algorithm³⁷ was developed during World War II for AOA geolocation using a flat (Cartesian) geometry, illustrated in Figure 47-20.

The LOB from a receiving site has a bearing angle β from north (y-axis). Using an arbitrary coordinate system, the LOB is also described by d , the perpendicular distance to the coordinate system origin; m is the LOB miss distance from point $(p_x, p_y)^T$.

The miss distances from a set of K lines of bearing with common Cartesian origin created by some number of DF sites is written in matrix form as

$$\mathbf{m} = \mathbf{A}\mathbf{p} + \mathbf{d} \quad (47-47)$$

where \mathbf{m} is the miss vector, \mathbf{A} is a transformation matrix, \mathbf{p} is the position vector, and \mathbf{d} is the LOB origin offset parameter:

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_K \end{bmatrix}, \mathbf{A} = \begin{bmatrix} -\cos(\beta_1) & \sin(\beta_1) \\ -\cos(\beta_2) & \sin(\beta_2) \\ \vdots & \vdots \\ -\cos(\beta_K) & \sin(\beta_K) \end{bmatrix}, \mathbf{p} = \begin{bmatrix} p_x \\ p_y \end{bmatrix}, \text{ and } \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix}$$

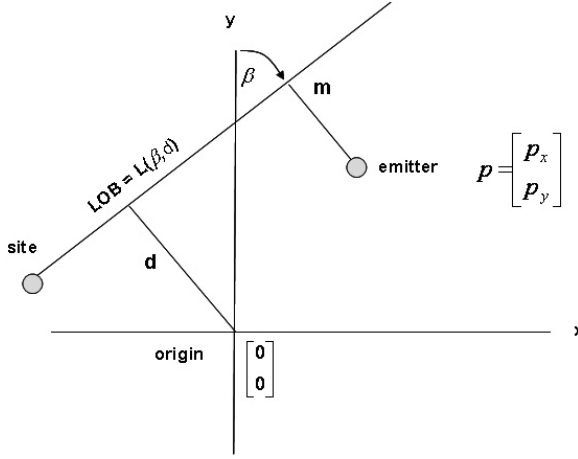


FIGURE 47-20 Stansfield geolocation geometry

The goal is to find a position $\hat{\mathbf{p}}$ that minimizes $\mathbf{m}^T \mathbf{m} = \sum_1^K m_i^2$, which can be done in an iterative approach using the gradient form of Eq. 47-47. The iterative improvement in position is

$$d\mathbf{p} = \text{inv}(\mathbf{A}^T \mathbf{A}) \cdot (\mathbf{A}^T \mathbf{m}) \quad (47-48)$$

and

$$\hat{\mathbf{p}}_{\text{new}} = \hat{\mathbf{p}}_{\text{old}} + d\mathbf{p} \quad (47-49)$$

Inspection of the components of Eq. 47-48 reveals that the $\mathbf{A}^T \mathbf{A}$ and $\mathbf{A}^T \mathbf{m}$ matrices are the classic unweighted least squares normal equations. These should be used on the first iteration from an arbitrary starting point \mathbf{p}_0 . Subsequent iterations should use a weighted version of $\mathbf{A}^T \mathbf{A}$ and $\mathbf{A}^T \mathbf{m}$ that account for the LOB a priori DF variance σ_{LOB}^2 and distance r_i from the i th bearing site to emitter. Using a covariance matrix \mathbf{R} whose diagonal elements contain the squared weights $w_i^2 = 1 / (\sigma_{\text{LOB}}^2 r_i^2)$, Eq. 47-48 is modified to be

$$d\mathbf{p} = \text{inv}(\mathbf{A}^T \mathbf{R} \mathbf{A}) \cdot (\mathbf{A}^T \mathbf{R} \mathbf{m}) \quad (47-50)$$

The weighted normal equations are written as

$$(\mathbf{A}^T \mathbf{R} \mathbf{A}) = \begin{bmatrix} \sum_{i=1}^K \frac{\cos^2(\beta_i)}{\sigma_{\text{LOB}}^2 r_i^2} & -\sum_{i=1}^K \frac{\cos(\beta_i) \sin(\beta_i)}{\sigma_{\text{LOB}}^2 r_i^2} \\ -\sum_{i=1}^K \frac{\cos(\beta_i) \sin(\beta_i)}{\sigma_{\text{LOB}}^2 r_i^2} & \sum_{i=1}^K \frac{\sin^2(\beta_i)}{\sigma_{\text{LOB}}^2 r_i^2} \end{bmatrix} \quad (47-51)$$

TABLE 47-2 Ellipse Confidence Factor

Ellipse Confidence Factor	Scaling Factor (jo)
68%	1
80%	1.79
90%	2.15
95%	2.45

and

$$(\mathbf{A}^T \mathbf{R} \mathbf{m}) = \begin{bmatrix} -\sum_{i=1}^K \frac{m_i \cos(\beta_i)}{\sigma_{\text{LOB}i}^2} \\ \sum_{i=1}^K \frac{m_i \sin(\beta_i)}{\sigma_{\text{LOB}i}^2} \end{bmatrix} \quad (47-52)$$

Convergence to $\hat{\mathbf{P}}$ requires only three or four iterations from an arbitrary starting point, making the Stansfield geolocation algorithm fairly efficient. It is customary to estimate a confidence containment ellipse about the location $\hat{\mathbf{P}}$, which can be determined from the eigenvalues of the $\text{inv}(\mathbf{A}^T \mathbf{R} \mathbf{A})$ matrix. Letting the inverse matrix be noted as \mathbf{Q} , the eigenvalues are found by solving the determinate $|\lambda \mathbf{I} - \mathbf{Q}| = 0$. Two eigenvalues result, designated λ_+ (maximum) and λ_- (minimum). The confidence ellipse is scaled by the number of LOBs, K , and by a scaling factor, jo , which establishes the statistical confidence level (see Table 47-2).

The confidence ellipse is then described by the semi-major and semi-minor axes:

$$s \text{ maj} = jo\lambda_+ \sqrt{\frac{K}{K-1}} \text{ and } s \text{ min} = jo\lambda_- \sqrt{\frac{K}{K-1}} \quad (47-53)$$

The unambiguous ellipse orientation (see Table 47-3) requires testing of the terms of the \mathbf{Q} matrix.

A Monte Carlo simulation of LOBs from three bearing sites illustrates the Stansfield algorithm. LOBs with 2° rms error were generated from three sites, producing a best estimate position and 95 percent confidence containment ellipse (see Figure 47-21a). To evaluate 300 trials, the confidence ellipse, normally centered on each estimated position, is shifted to the true emitter location, and allows comparison of the theoretical 68%, 80%, and 95% confidence ellipse containment with the trial estimated positions (see Figure 47-21b).

TABLE 47-3 Confidence Ellipse Orientation

Compare Diagonal Terms	Orientation
$Q_{1,1} \leq Q_{2,2}$	$90 - \frac{1}{2} \text{atan} 2 \left(\frac{y \text{ term}}{x \text{ term}} \right)$
$Q_{1,1} > Q_{2,2}$	$\frac{1}{2} \text{atan} 2 \left(\frac{y \text{ term}}{x \text{ term}} \right)$
where	$y \text{ term} = Q_{1,2} + Q_{2,1}$ $x \text{ term} = Q_{1,1} - Q_{2,2}$

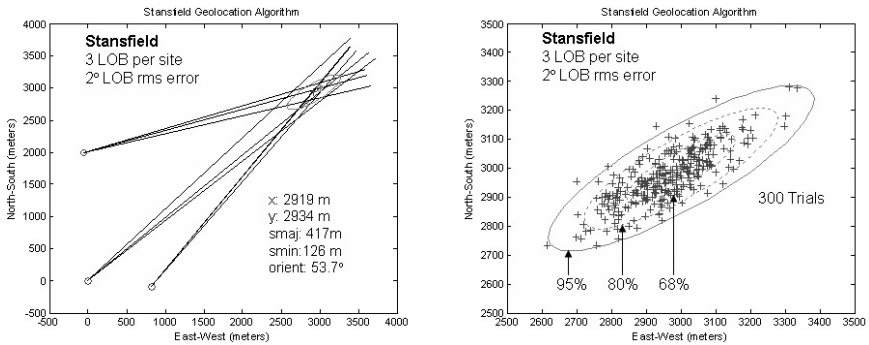


FIGURE 47-21 (a) Stansfield LOBs, geolocation, and 95% confidence ellipse and (b) 300 trials with 68%, 80%, and 95% confidence ellipse centered on the true emitter

Wangness AOA Geolocation Algorithm

In 1973 Dennis Wangness proposed an elegant solution to geolocation on a spherical earth.³⁸ It has the advantage of being a fast, non-iterative eigenvector approach to produce both geolocation and a confidence containment ellipse. Extending the spherical solution onto a geoid allows the algorithm to be WGS-84 compliant.

Wangness recognized that any LOB describes a great circle around the earth and is anchored through the bearing site. These great circles are described by perpendicular normal vectors, \mathbf{n} , which collectively produce an eigenvector solution for best estimate of position and error ellipse.

First, the LOB normal vectors are described in terms of the Earth Centered Earth Fixed (ECEF) coordinate system and are collected into transformation matrix \mathbf{A} for K LOBs:

$$\mathbf{A} = \begin{bmatrix} nx_1 & ny_1 & nz_1 \\ \vdots & \vdots & \vdots \\ nx_K & ny_K & nz_K \end{bmatrix} = [\mathbf{n}_1 \quad \cdots \quad \mathbf{n}_i \quad \cdots \quad \mathbf{n}_K]^T \quad (47-54)$$

where $\mathbf{n}_i = \begin{bmatrix} nx_i \\ ny_i \\ nz_i \end{bmatrix}$ is the i th LOB normal vector in ECEF coordinates.

Once the $\mathbf{A}^T\mathbf{A}$ matrix is formed, the Wangness algorithm solves for the three eigenvalues and associated eigenvectors using $|\lambda\mathbf{I} - \mathbf{A}^T\mathbf{A}| = 0$. The normalized eigenvector associated with the smallest eigenvalue is a unit vector that points toward the best estimate of emitter position. This unit vector can be transformed into spherical latitude and longitude and further transformed into geodetic latitude and longitude using a standard geoid such as WGS-84.

Although the eigen solution is not iterative, once an approximate emitter position is found, a weighted solution using the covariance error matrix $\mathbf{R} = \text{diag}(1/r_i^2)$, where r_i is the arc distance between the i th-LOB bearing site and the estimated location. The $\mathbf{A}^T\mathbf{A}$ matrix now becomes the weighted matrix $\mathbf{A}^T\mathbf{R}\mathbf{A}$.

The three eigenvalues of $|\lambda \mathbf{I} - \mathbf{A}^T \mathbf{R} \mathbf{A}| = 0$ are sorted smallest to largest (λ_1 , λ_2 , and λ_3) to provide the confidence ellipse equations:

$$s_{\text{maj}} = \frac{a_e \cdot jo}{\sqrt{K-2}} \cdot \sqrt{\frac{\lambda_1}{\lambda_2}} \quad \text{and} \quad s_{\text{min}} = \frac{a_e \cdot jo}{\sqrt{K-2}} \cdot \sqrt{\frac{\lambda_1}{\lambda_3}} \quad (47-55)$$

where K is the number of LOBs, jo is the ellipse scaling factor for confidence of containment, and a_e is the earth's radius to convert arc distances into meters. Ellipse orientation is derived from the arc cosine of the dot product formed by the eigenvector associated with λ_2 (the eigenvalue associated with semi-major axis) and a unit vector at the estimated location that points east.

A similar Monte Carlo simulation was run for the Wangness algorithm with the same LOB error and site geometry of the Stansfield simulation. LOB, best position, and confidence ellipse were computed (see Figure 47-22a) and 300 trials were made (see Figure 47-22b). Again, the confidence ellipse was shifted to the true emitter location, for comparison of the theoretical 68%, 80%, and 95% confidence containment ellipses.

TDOA Geolocation Algorithm

When the LORAN hyperbolic location system was first developed by MIT in 1942, it was computationally intensive and highly classified.³⁹ Now TDOA geolocation can be done with small radio tuners and inexpensive computer chips. The algorithm presented here is consistent with the two-dimensional Stansfield approach, but can be easily extended to a three-dimensional solution on a spherical earth or WGS-84 geoid.^{40,41}

The time difference of a signal arriving at two sites s_1 and s_2 from an emitter at s_o can be written in terms of the two distances D_1 and D_2 between the sites and emitter, scaled by the speed of light, c :

$$\text{TDOA}_{ik} = \frac{D_i - D_k}{c} \quad (47-56)$$

where D_i is the distance between site s_i and emitter at s_o

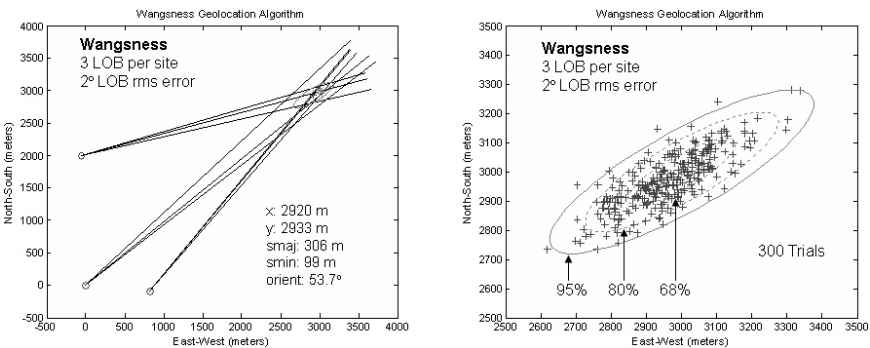


FIGURE 47-22 (a) Wangness LOBs, geolocation, and 95% confidence ellipse and (b) 300 trials with 68%, 80%, and 95% confidence ellipse centered on the true emitter

$$D_i = |\mathbf{s}_i - \mathbf{s}_o| = \{(x_i - x_o)^2 + (y_i - y_o)^2\}^{\frac{1}{2}} \quad (47-57)$$

The hyperbolic TDOA solution of location is written in an iterative, gradient form similar to the Stansfield approach of Eqs. 47-48 and 47-49:

$$d\mathbf{s}_o = (\mathbf{A}^T \mathbf{A}) \cdot d(\mathbf{TDOA}) \quad (47-58)$$

and

$$\hat{\mathbf{s}}_{\text{new}} = \hat{\mathbf{s}}_o + d\mathbf{s}_o \quad (47-59)$$

The matrix $d(\mathbf{TDOA})$ is a $K \times 1$ vector of K observed minus calculated TDOA values from the i - k site pair, where the calculated TDOA _{ik} and its derivative are based on the current estimate of emitter position $\hat{\mathbf{s}}_o$. The $\mathbf{A}^T \mathbf{A}$ matrix is a $K \times 2$ transformation matrix of partial derivatives of TDOA _{ik} with respect to $\hat{\mathbf{s}}_o$. The fundamental $d(\mathbf{TDOA})$ vector and \mathbf{A} matrix are written as

$$d(\mathbf{TDOA}) = \begin{bmatrix} \text{TDOA}_{ik}(\text{measured})_1 - \text{TDOA}_{ik}(\text{calculated})_1 \\ \text{TDOA}_{ik}(\text{measured})_2 - \text{TDOA}_{ik}(\text{calculated})_2 \\ \vdots \\ \text{TDOA}_{ik}(\text{measured})_K - \text{TDOA}_{ik}(\text{calculated})_K \end{bmatrix} \quad (47-60)$$

and

$$\mathbf{A} = \begin{bmatrix} \left\{ \frac{\partial \text{TDOA}_{ik}}{\partial x_o} dx_o \right\}_1 & \left\{ \frac{\partial \text{TDOA}_{ik}}{\partial y_o} dy_o \right\}_1 \\ \left\{ \frac{\partial \text{TDOA}_{ik}}{\partial x_o} dx_o \right\}_2 & \left\{ \frac{\partial \text{TDOA}_{ik}}{\partial y_o} dy_o \right\}_2 \\ \vdots & \vdots \\ \left\{ \frac{\partial \text{TDOA}_{ik}}{\partial x_o} dx_o \right\}_K & \left\{ \frac{\partial \text{TDOA}_{ik}}{\partial y_o} dy_o \right\}_K \end{bmatrix} \quad (47-61)$$

The partial differential of TDOA _{ik} with respect to the current estimate of emitter position is further expanded using the definitions of time delay from Eqs. 47-56 and 47-57:

$$\frac{\partial \text{TDOA}_{ik}}{\partial x_o} = \frac{1}{c} \left(\frac{\partial D_i}{\partial x_o} - \frac{\partial D_k}{\partial x_o} \right) = \frac{1}{c} \left(\frac{(x_o - x_i)}{D_i} - \frac{(x_o - x_k)}{D_k} \right) \quad (47-62)$$

and

$$\frac{\partial \text{TDOA}_{ik}}{\partial y_o} = \frac{1}{c} \left(\frac{\partial D_i}{\partial y_o} - \frac{\partial D_k}{\partial y_o} \right) = \frac{1}{c} \left(\frac{(y_o - y_i)}{D_i} - \frac{(y_o - y_k)}{D_k} \right) \quad (47-63)$$

At least two different site pairs are required for geolocation convergence, and care must be taken to avoid ambiguities, since, unlike LOB observations, the hyperbolic TDOA isochrones have left-right symmetry about a line between the measuring i - k site pair.

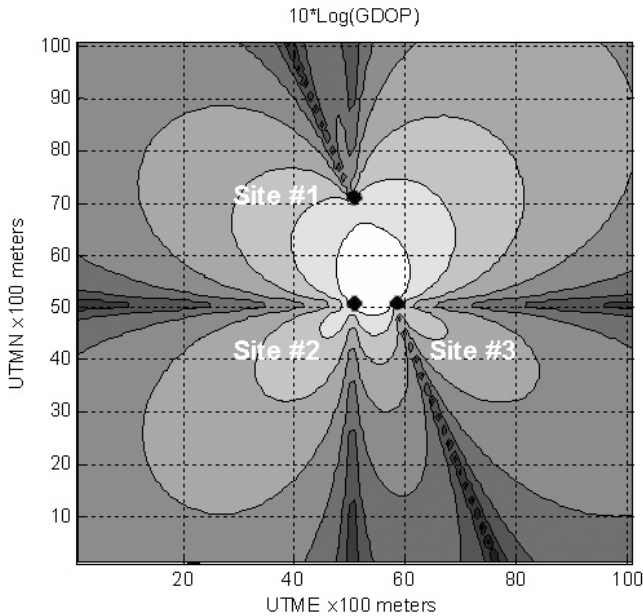


FIGURE 47-23 GDOP for TDOA geolocation from three sites. Lightest color indicates lowest GDOP and area of best geolocation.

A useful concept in evaluating the geolocation geometry is a quantity called the Geometric Dilution of Precision (GDOP), defined as

$$\text{GDOP} = \text{trace}(\text{inv}(\mathbf{A}^T \mathbf{A})) \quad (47-64)$$

A low value of GDOP indicates good geometry for the geolocation solution, while a high GDOP indicates poor geometry (see Figure 47-23).

A Monte Carlo simulation with a geometry similar to Stansfield and Wangsness was run for TDOA geolocation (see Figure 47-24). Each TDOA has $\sigma_{\text{TDOA}} = 50\text{-ns}$ rms error.

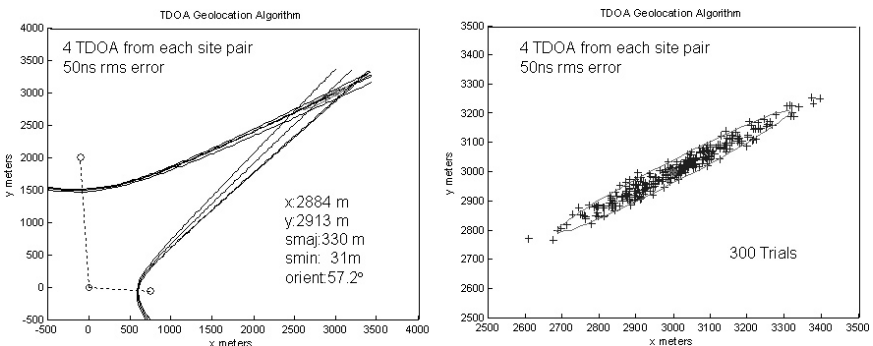


FIGURE 47-24 (a) TDOA hyperbolic isochrones, geolocation, and 95% confidence ellipse and (b) 300 trials with 95% confidence ellipse centered on the true emitter

Although the TDOA algorithm estimates position entirely in the time domain, the simulation also drew hyperbolic isochrones for comparison with the LOB geolocation approach (compare Figure 47-24a with Figures 47-21a and 47-22b). Four TDOA measures were made from each site pair (shown as dotted lines connecting the sites). The calculation of the confidence error ellipse closely parallels that of the Stansfield algorithm described by Eq. 47-53 and Table 47-3. In Figure 47-24b, the confidence ellipse is shifted to the true emitter location to examine the 95% containment of 300 geolocation trials. The TDOA geolocation error is distributed very differently from the LOB geolocation error (compare Figure 47-24b with Figures 47-21b and 47-22b).

REFERENCES

1. F.E. Terman, *Radio Engineering* (New York: McGraw-Hill, 1947): 817.
2. National Telecommunications and Information Administration (NITA), *Manual of Regulations and Procedures for Federal Radio Frequency Management*, Section 6.1.1 (Jan. 1993): 6–12.
3. F. B. Gross, *Smart Antennas for Wireless Communications with MatLab* (New York: McGraw-Hill, 2005).
4. S. C. Hooper, "Navy History—Radio, Radar, and Sonar," Recordings, Office of Naval History, Washington, D.C., 31R76. [Quoted references may be found in L. S. Howeth, *History of Communications-Electronics in the United States Navy*, Chapter XII, "The Radio Direction Finder": <http://earlyradiohistory.us/1963hw22.htm>.]
5. L. S. Howeth, *History of Communications-Electronics in the United States Navy* (Washington, D.C.: United States Government Printing Office, 1963): <http://earlyradiohistory.us/1963hw.htm>.
6. F. H. Engel and F. W. Dunmore, "A Directive Type of Radio Beacon and Its Application to Navigation," *Scientific Papers Bureau of Standards*, vol. 19 (1923): 281–295.
7. G. Springer, "Radio Systems in the Early A6M Zero," http://www.jaircraft.com/research/gregspringer/radios/radio_systems.htm.
8. "History or Radio Research at Ditton Park," <http://www.dittonpark-archive.rl.ac.uk/histTime.html>.
9. D. Kahn, *The Codebreakers* (New York: Macmillan Co., 1969): 8/504.
10. V.A. Kotelnikov, "Determination of the Elements of the Orbit of a Satellite Using the Doppler Effect," *Radiotekhnika and Elektronika*, vol. 7 (July 1958): 873–991.
11. C. Soanes, "Orbital Navigation Systems—Present and Future Systems," Dynamic Positioning Maritime Technology Society Conference, London, Sept. 28–30, 2004. [Paper available at http://dynamic-positioning.com/dp2004/sensors_soanes.pdf.]
12. P. Daly and G. E. Perry, "Recent Developments with the Soviet Union's VHF Satellite Navigation System," *Space Communications and Broadcasting*, vol. 4 (1986): 51–61.
13. F. W. Lehan and G. L. Brown, Discovery and Location System (Nov. 6, 1962): U.S. Pat. 3,063,048.
14. <http://www.cospas-sarsat.org/> and <http://www.sarsat.noaa.gov/>.
15. R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. on Antennas and Propagation*, vol. AP-34(3) (1986): 276–280.
16. R. Roy and T. Kailath, "ESPRIT—Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. ASSP-37 (July 1989) 984–995.
17. A. Moffet, "Minimum-Redundancy Linear Arrays," *IEEE Trans AP*, vol. 16(2) (1968): 172–175.
18. S. Stein, "Algorithms for Ambiguity Function Processing," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-29 (June 1981): 588–599.
19. C. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. ASSP-24 (Aug. 1976).
20. P. J. D. Gething, *Radio Direction Finding and Superresolution*, 2nd Ed. (London: Peter Peregrinus Ltd. on behalf of IEE, 1991): 94–113.

21. C. W. Earp and R. M. Godfrey, "Radio Direction Finding by Measurement of the Cyclical Difference of Phase," *J. IEE (London)*, part IIIA, vol. 94 (March 1947): 705.
22. F. Adcock, Improvements in Means for Determining the Direction of a Distant Source of Electro-Magnetic Radiation (1919): British Pat. 130,490.
23. R. Poisel, *Introduction to Communication Electronic Warfare Systems* (Boston: Artech House, 2002): 352–359.
24. D. H. Johnson, "The Application of Spectral Estimation Methods to Bearing Estimation Problems," *Proc. IEEE*, vol. 70(9) (Sept. 1982): 1018–1028.
25. J. P. Burg, "Maximum Entropy Spectrum Analysis," Ph.D. dissertation, Dept. of Geophysics, Stanford University, Stanford, CA, 1975.
26. V. F. Pisarenko, "The Retrieval of Harmonics from a Covariance Function," *Geophys. J. Royal Astronomical Soc.*, vol. 33 (1973): 347–366.
27. R. Kumaresan and D. W. Tufts, "Estimating the Angles of Arrival of Multiple Plane Waves," *IEEE Trans. Aerospace Electron. Sys.*, vol. AES-19 (Jan. 1983) 134–139.
28. I. Ziskind and M. Wax, "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. ASSP-36 (Oct. 1988): 1553–1560.
29. E. Anderson et al., *LAPACK Linear Algebra Package*, <http://math.nist.gov/lapack++/>.
30. J. Volder, "The CORDIC Computing Technique," *IRE Trans. Computers*, vol. EC-8 (Sept. 1959): 330–334.
31. A. Ahmedsaid, A. Amira, and A. Bouridane, "Improved SVD Systolic Array and Implementation on FPGA," *Proc. 2003 IEEE Inter. Conf. on Field-Programmable Technology (FPT)* (Dec. 15–17, 2003): 35–42.
32. G. W. Stewart, *Matrix Algorithms Vol. II: Eigensystems* (Philadelphia: Society for Industrial and Applied Mathematics, 2001): 56–70.
33. H.L. Van Trees, *Optimum Array Processing* (New York: Wiley Interscience, 2002): 925–932.
34. H.L. Van Trees, *Optimum Array Processing*: 827–845.
35. H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Automatic Control*, vol. AC-19 (June 1974): 716–723.
36. M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria," *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. ASSP-33 (April 1985): 387–392.
37. R. G. Stansfield, "Statistical Theory of DF Fixing," *J. IEE (London)*, part IIIA, vol. 94 (1947): 762–770.
38. D. Wangsness, "A New Method of Position Estimation Using Bearing Measurements," *IEEE Trans. Aerospace and Electronic Sys.* (Nov. 1973): 959–960.
39. D. A. Grier, *When Computers Were Human* (Princeton, NJ: Princeton University Press, 2005): 249–250.
40. A. D. Stewart, "Comparing Time-Based and Hybrid Time-Based/Frequency Based Multi-Platform Geolocation Systems," Thesis, Naval Postgraduate School, Monterey, CA, Sept. 1997.
41. H. C. Schau and A. Z. Robinson, "Passive Source Location Employing Spherical Surfaces from Time-of-Arrival Differences," *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. ASSP-35 (Aug. 1987) 1223–1225.

Chapter 48

ESM and ECM Antennas

Nikolaos K. Uzunoglu
George Geroulis

National Technical University of Athens

CONTENTS

48.1	INTRODUCTION.....	48-2
48.2	COMMUNICATION ESM AND ECM SYSTEM ANTENNAS.....	48-3
48.3	RADAR ESM AND ECM SYSTEM ANTENNAS	48-9

48.1 INTRODUCTION

ESM and ECM Systems

The use of electronic countermeasures (ECM) against opponents' communications systems started with the use of wireless communications in military operations at the beginning of the 20th century. Extensive electronic monitoring/interception of nonfriendly signals and countermeasures against both communications links (primarily high frequency [HF] at that time) and newly developed radar systems began during World War II (1939–45). The use of the electromagnetic spectrum in military operations was defined as “electronic warfare.” Soon these activities were classified as electronic support measures (ESM). ESM consisted of passive measurements of emitted signals by nonfriendly emitters, and ECM systems, which included the emission of signals to confuse or deceive the communications and radar sensor systems of opposing forces.

Extensive use of passive ECM techniques, such as chaff (a large number of dipole elements spread into the air), began toward the end of WWII. Electronic counter-counter measure (ECCM) techniques are aimed to overcome or decrease the ECM threats of friendly communications and sensor systems. To this day, ECM and ECCM techniques are in a continuous competition.

The design and development of antennas intended to serve the needs of ESM and ECM systems poses unique requirements, as compared to communications or radar antennas, such as

- Ultrawide spectral bandwidths
- Low weight and size
- Full polarization control
- Very high isolation between transmit and receive in collocated antennas

The last requirement is of particular significance, since the ability to receive weak signals while transmitting countermeasure (jamming) signals is of primary importance. Achievement of this goal provides superior ECM abilities. Indeed, the capability to regenerate threat signals improves the efficiency of ECM systems to a high degree.

General Properties of ESM and ECM Antennas

In treating the ESM/ECM antennas the spectral region of interest has primary importance. In tactical communications systems the frequency region 1–2500 MHz (HF, VHF) is used predominantly, while radars use microwave (1–30 GHz) and millimeter wavelength (30–100 GHz) spectral regions. The wideband antenna technologies, which started to appear during the mid-1950s, are widely used in both ESM and ECM applications concerning both communications and radar systems. The extensive use of multiple communications networks at the same geographical place and jamming-resistant wideband systems forces the use of wideband systems, thus requiring ultra wideband (UWB) ESM and ECM antennas. Furthermore, especially in low-frequency ESM and ECM systems, the relatively large size of antennas necessitates the use of the same antenna aperture for various needs such as signal interception and emission at various frequency bands almost simultaneously.

The development of monolithic integrated circuits (based on GaAs or Si technologies) after the 1980s enabled the use of phased-array principles in ESM and ECM systems.

The availability of true time-delay, digitally controlled, two-port networks allowed the development of wideband phased-array antennas, which provide superior operational characteristics such as the following:

- Highly dynamic beam agility in one or two planes
- Ability to modify outgoing wavefront to effect monopulse radars
- Use of nulling at the direction of arrival energy jamming signals without significant degradation of radiation pattern at other angles
- Highly effective irradiated power toward threats

In specifying and designing ESM and ECM antennas, the standard quantities used in general antenna theory are used, such as directivity (D), gain (G), bandwidth (BW), and radiation patterns at various levels.

In ECM and ESM system antennas, the polarization properties of antennas are of paramount importance to performance. The polarization behavior of ECM systems, and to some extent ESM systems, has essential effects on the ability of radars to resist jamming and deception phenomena (ECCM). Furthermore, the level of isolation between receiver and transmitter antennas in an ECM system, either for communications or radar, is of fundamental importance because it affects to a significant degree the countermeasure capabilities.

In this chapter, the ESM and ECM systems antennas are reviewed independently concerning the communications and radar systems. The presentation reflects the practical use of antennas in the corresponding applications based on generic concepts. In the case of communications systems, attention is focused on the low frequencies (1–2500 GHz), while for radar applications the microwave (1–30 GHz) and millimeter wavelength (30–100 GHz) bands are considered. In all the considered cases, the mechanical, environmental, and radar cross-section implications are also considered in practical terms. In both cases, the direction-finding antenna principles are also discussed and investigated based on various monopulse principles. At the end of the chapter, the important issue of transmit-receive antenna isolation in ECM systems is analyzed and general rules to achieve maximum isolation are suggested.

48.2 COMMUNICATION ESM AND ECM SYSTEM ANTENNAS

Mission of Communication ESM and ECM Systems

Communication ESM Systems The operational mission of communication ESM systems is to detect-receive incident, usually very weak, signals and convey them to sensitive wideband receivers. There are two fundamental operational missions of ESM systems: (a) acquisition and analysis of incoming signals and (b) measuring direction of arrival (direction finding, or DF) at the horizontal plane with respect to the Earth's surface. In the former case, signal parameters are measured without determining the angle of arrival in the horizontal plane.

Usually, omnidirectional antennas are utilized to receive signals without measurement of angle of arrival. Near the earth's surface, at the low radio frequencies (RF), 1–500 MHz, only vertical polarization is a matter of interest, while at higher frequencies, reception horizontal polarization should also be taken into account.

In the case of DF systems, monopulse antenna principles based on the simultaneous measurement and comparison of received signal amplitudes or phases is utilized (see Chapter 47). The DF systems are required to measure wideband signals such as the case of frequency-hopping and/or spread-spectrum modulated carriers, which pose serious difficulties in achieving azimuth angle measurement accuracies on the order of a few degrees. Furthermore, a dynamic range of at least 80 dB is required. The extensive use of digital signal processing techniques during recent years has enabled simultaneous measurement of DF and signals tracking of frequency-hopping and spread-spectrum signals. The combination of DF measurement techniques and use of signal processing techniques such as FFT and wavelet transform has provided the capability of efficient multiple signals processing and de-interleaving. Therefore, the modern trend is to have combined DF and signal acquisition-analysis ESM receivers, which require a single antenna for both missions.

Communication ECM Systems The primary role of an ECM antenna in communications applications is to direct the radiated countermeasure signals toward the enemy receivers to achieve degradation or complete loss of signal reception at the opponent's receiver units. This means that the azimuthal orientation of the victim receiver (or receivers) should be known with an accuracy of at least 10° in tactical communications using VHF (30–90 MHz) or UHF (200–2500 MHz) frequencies. In such cases, mobile ECM systems are usually used, placed inside a shelter for land-based systems. Ships or airplanes (manned or unmanned) are also used to a lesser degree. A significant difficulty arises because of the large antennas needed for ECM tactical systems required to cover the low VHF region, corresponding to this wavelength $\lambda = 3\text{--}10\text{ m}$.

Concerning HF communication ECM systems, usually land-based systems, antenna arrays with wideband properties are used. In this case, large-size fixed-array installations are used. In mobile, land-based HF ECM systems, usually omnidirectional or medium-directivity antennas with wideband characteristics are used.

Omnidirectional Antennas To achieve wideband reception of signals omnidirectionally, the most commonly used antennas are based on conical type structures, as shown in Figure 48-1, which also shows the corresponding critical design parameters.

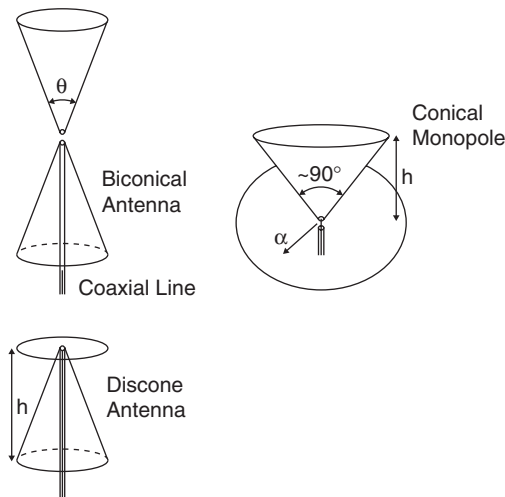


FIGURE 48-1 Family of conical antennas

In practice, the geometric size restrictions pose difficulties in achieving low-frequency coverage. The use of conical monopole or disccone antennas usually is preferable since the required antenna heights are $h = \sqrt{2} \lambda/8$ and $h = \lambda/8$, respectively. In the case of VHF and UHF frequencies, to reduce the antenna weight and wind load, radial conductors are used as shown in Figure 48-2. The number of radials should be at least 16 peripherally to achieve sufficiently dense current distributions.

A significant issue in omnidirectional antennas is the ability to receive and transmit both polarizations simultaneously, especially in the case of microwave signals and radar signals. The simplest approach is to polarize the incident-outgoing waves by using a grid of parallel wires, as shown in Figure 48-3. The distance between the two parallel running wires should be less than $\lambda_M/20$, where λ_M is the maximum wavelength of the radiation. Usually, printed strips of approximately $\lambda_M/100$ on a low-permittivity substrate are used. The drawback of this method is the 3 dB insertion loss, concerning either vertical or horizontal polarization.

In the case of HF (wavelength $\lambda = 100\text{--}10\text{ m}$) and VHF ($\lambda = 10\text{--}1\text{ m}$) frequencies, reception of both polarizations rarely is needed. Therefore, in the case of UHF ($\lambda = 1\text{--}0.1\text{ m}$) and SHF, EHF ($\lambda = 10\text{--}1\text{ cm}$) frequencies, in order to receive both polarizations with an omnidirectional radiation pattern, one can use independent antennas and then use either independent analog summation of the two signals or independent signal processing channels.

Figure 48-4 depicts two techniques to achieve reception and transmission for both polarization antenna systems. In Figure 48-4a the printed and slot triangular butterfly antennas

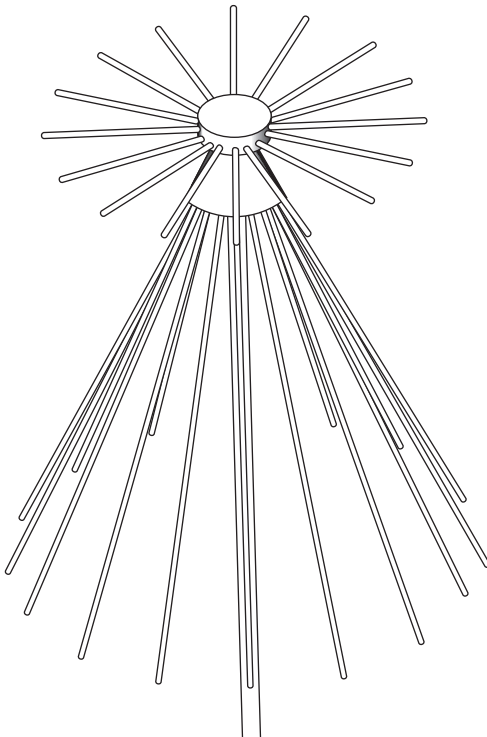


FIGURE 48-2 Disccone antenna made of radial conductors

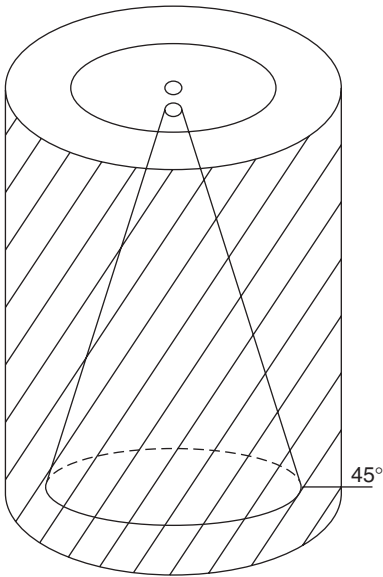


FIGURE 48-3 Polarized discone conical antenna

are placed collinearly to simultaneously radiate both polarizations. Another example is shown in Figure 48-4*b* where orthogonal conical dipoles are used to receive at all possible polarizations and incident angle signals.

Directional Antennas The antenna type that has been used, and still is widely used, in ESM and ECM communications systems is the logarithmic periodic antenna (LPA), which

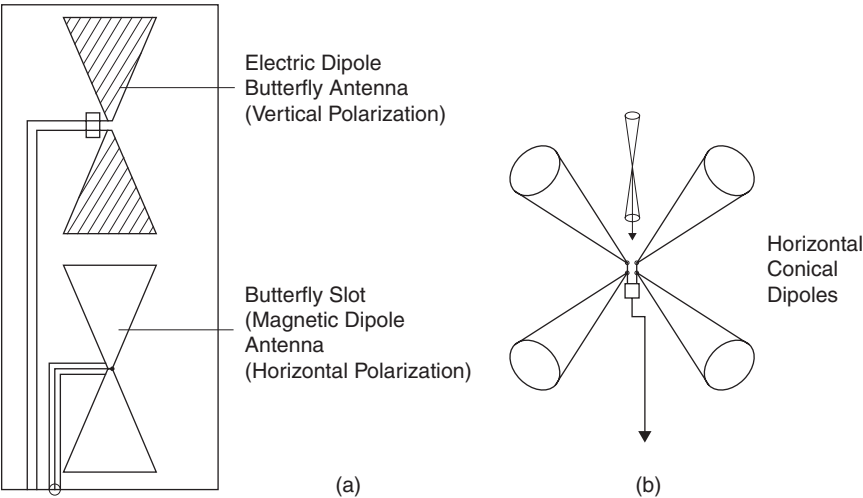


FIGURE 48-4 Combination of electric-magnetic wideband dipoles

was invented by Raymond H. Du Hamel.¹⁻³ The following are the critical design parameters in LPAs:

- Maximum and minimum operation frequency
- The required directivity and gain
- Polarization characteristics

The basic structure of an LPA is shown in Figure 48-5. The basic design equations of the structure of an LPA are also given. Although numerous variations of LPAs have been developed and used during the last 50 years, the structure with multiple dipole elements has been the most commonly used ECM and ESM wideband antenna in practice. LPAs could provide antenna gains on the order of 7-10 dBi depending on the number of elements (N) and apex angle (α). If more directivity is required, use of multiple LPAs proved to be a useful technique, such as shown in Figure 48-6, where the two LPAs are placed parallel with their axis forming the angle β and the distance between the largest dipole is equal to the largest dipole (half wavelength $\lambda_M/2$ at the largest frequency).

$$\frac{l_n + 1}{l_n} = \frac{h_n + 1}{h_n} = \tau$$

$$\tan(\alpha/2) = \frac{l_n - l_{n+1}}{h_n}$$

Maximum operation wavelength = $\lambda_M = 4l_1$ (largest dipole double length)

$$\sin\left(\frac{\beta}{2}\right) = \frac{\lambda_M/4}{L}$$

where

$$L = \sum_{n=1}^N h_n$$

N = number of elements

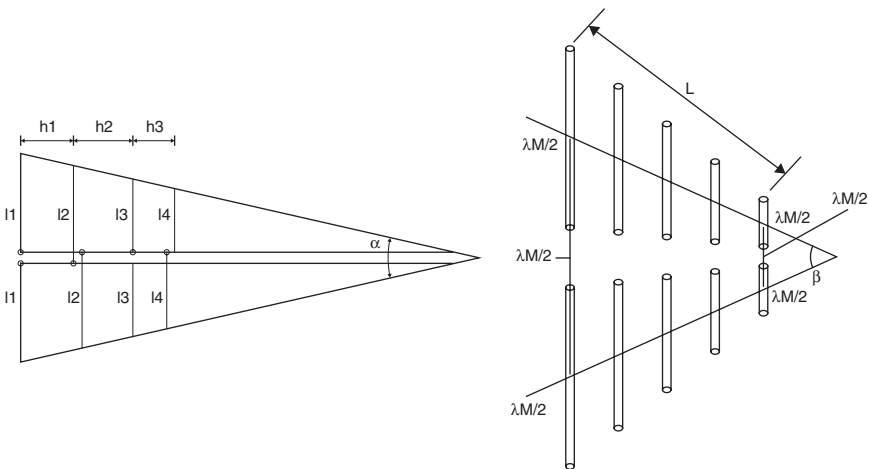


FIGURE 48-5 Array of log-periodic arrays

Communication Systems Direction Finding Antennas The simultaneous reception of signals with different radiation pattern antennas—a principle known as *monopulse*—is used to measure the angle of incidence of communications and radar signals. Usually the azimuth angles of the incident signals are measured. In this case, an even number of wideband antennas, such as $n = 4, 6, 8$, etc. directive antennas, are used, as shown in Figure 48-6. For low frequency communications systems, only vertical polarization reception is typically required, simplifying the antenna element construction. A commonly used wide antenna type is the transverse electromagnetic mode (TEM) horn antenna, shown in Figure 48-6a, in which log-periodic elements are also used (see Figure 48-6b). Analog signal processing techniques consisting of hybrid circuits (sum and difference) and down converter units are employed to drive the A/D conversion circuits for direction finding (see Chapter 47) using the two signals $X(\varphi)$ and $Y(\varphi)$. Assuming each individual antenna in Figure 48-6 has a radiation pattern function described by an even function $F(\varphi) = F(-\varphi)$ with respect to the $\varphi = 0$ axis, the $X(\varphi)$ and $Y(\varphi)$ functions are computed as follows:

$$X(\varphi) = F(\varphi) - F(\varphi - 180^\circ)$$

$$Y(\varphi) = F(\varphi - 90^\circ) - F(\varphi - 270^\circ)$$

By comparing the two signals $X(\varphi)$, $Y(\varphi)$ we can then obtain the φ angle as a function of the X and Y signals on the assumption of a monotonically decreasing function. Calibration techniques should be used to account for any frequency dependence of the $F(\varphi)$ radiation pattern function. Increased bandwidths and conversion rates of the A/D converters provides opportunity for eliminating the necessity of using analog processing units to obtain the $X(\varphi)$ and $Y(\varphi)$ signals by digital processing.

Instead of using directive antenna elements, many systems use a more simplified technique, known as the Watson-Watt technique, shown in Figure 48-6c, where four dipole elements are used to obtain the $X(\varphi)$ and $Y(\varphi)$ functions. The distance between the two antipodal dipole elements should be less than half a wavelength at the maximum operation frequency. This condition guarantees the unique dependence of φ to X and Y voltages. To cover wide bandwidths, three or four sets of collinearly placed quadrant dipole or monopole elements are used along the vertical axis.

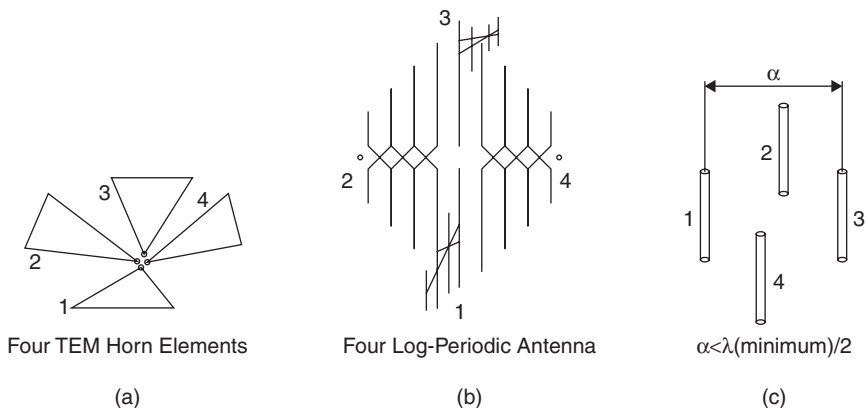


FIGURE 48-6 Wideband direction-finding antennas

48.3 RADAR ESM AND ECM SYSTEM ANTENNAS

Mission of Radar ESM and ECM Systems

Detection of radar signals with high-sensitivity receivers provides the opportunity of sensing the presence of radar emitters at large distances. The introduction of low probability of intercept (LPI) radars, which use very low peak power levels (few watts) with coded signal emissions, poses challenges in detecting such emitters with conventional noncoherent ESM receivers. Traditionally, because of the high peak power of pulsed radars, crystal video principle “envelope detection” receivers have been utilized. Furthermore, it is highly important to measure the direction of arrival of signals at the azimuth plane. Both direction and direction finding operations are always carried out simultaneously at many octave bandwidths (1–100 GHz). The required dynamic range of the receivers should be even more, at least up to 100 dB.

In ECM systems, depending on the platform to be protected (self) or the required effective radiated power to be emitted toward the victim radar, antennas ranging from moderate gain (10 dBi) to high gain (30 dBi) are used. In designing ECM antennas, the operational requirements as well as installation restrictions should be taken into account. In case of radar systems, since high microwave frequencies usually are employed, isolation levels between receive-transmit antennas as high as 90 dB could be achieved, which increases the capability of ECM systems.

ESM-DF Antennas to Detect Radar Signals To cover multi-octave bandwidths, inherently wideband antennas such as conical dipole/monopoles, spirals, and spiral conical structures are utilized. In traditional ESM-DF receivers, an omnidirectional antenna is separately used to drive an analysis channel while six to eight cavity-backed spiral antennas are used to measure the direction of arrival of the incoming signals. Figure 48-7 shows the architecture of an ESM-DF system (see also Chapter 47). To achieve wideband coverage (2–18 GHz or 0.5–40 GHz in many cases), direct detection receivers based on instantaneous frequency measurement correlators and amplitude detection have been used in the past. However, a serious drawback of these approaches has been the inability to detect continuous wave emitters and even more low probability intercept radars that incorporate encoded low power signals.

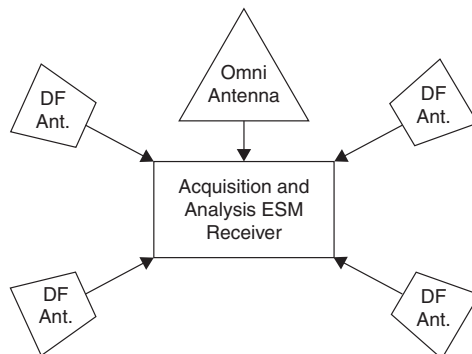


FIGURE 48-7 Radar signal ESM-DF system

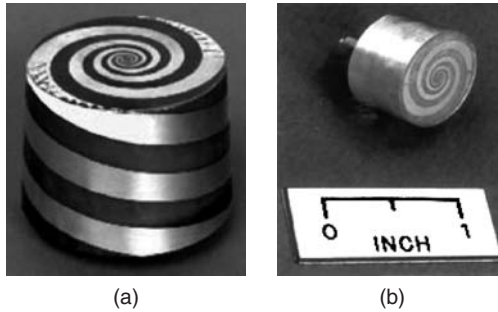


FIGURE 48-8 Spiral antennas used in DF antennas

Omnidirectional antennas with conical type structures, previously shown in Figure 48-1, are usually used to achieve wideband coverage. For the DF antennas, spiral antennas are commonly used, such as cavity-backed or conical spiral antennas. Figure 48-8 shows structures commonly used in direction finding spiral antennas.

In ECM systems against radars, in the case of medium-gain antennas, wideband horn-type antennas are used, as shown in Figure 48-9, where loading of conventional horn antennas by strips is used. To achieve higher gains, parabolic reflector antennas using wideband horns are utilized. Phased-array antennas using wideband properties are also used extensively to achieve high gains and fast control of radiation beams.

The development of high-speed analog to digital (A/D) converters achieving up to several Giga-Sample/second (GS/s) rates, can allow the development of fully coherent receivers using wideband down-converters prior to digitization using A/D converters operating at 1–2 GS/s. In these systems, either a spinning antenna or, again, a circular array of wideband antennas (i.e., cavity spiro-conical or even wideband horn antennas) is used. Figure 48-10 shows the architecture of this type of system. The use of a rotating passive reflector excludes the use of a rotational joint, which causes reliability problems and difficulties in achieving wideband properties.

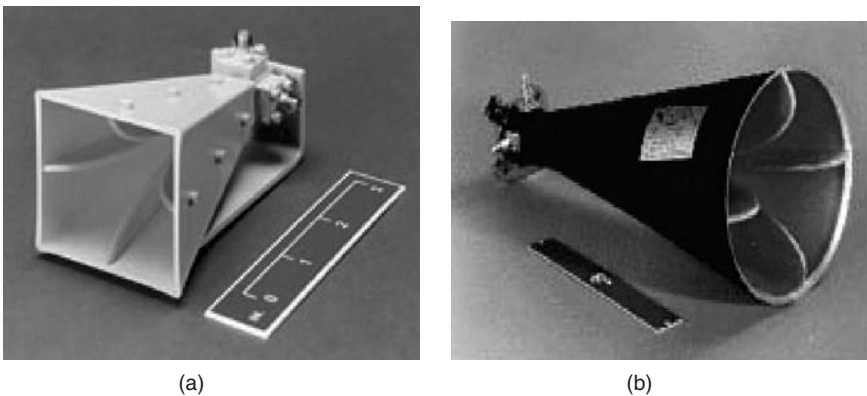


FIGURE 48-9 Wideband horn antennas

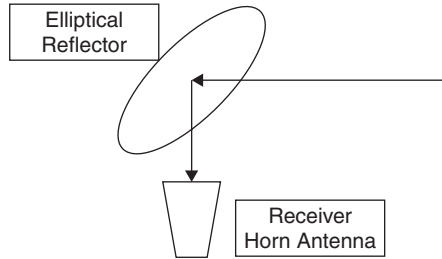


FIGURE 48-10 Spinning direction finding antenna

REFERENCES

1. R. H. Du Hamel, Frequency Independent Split-Beam Antennas, U.S. Pat. 2,977,597.
2. A.T. Mahnad, "A Tribute to a Great Inventor: Raymond H. Du Hamel 1921–1993," *IEEE Antennas and Propagation Magazine*, vol. 36, no. 3 (June 1994): 54–56.
3. G. A. Deschamps and J. D. Dyson, "The Logarithmic Spiral in a Single-Aperture Multimode Antenna System," *IEEE Trans. on Antennas and Propagat.*, vol. AP-19, no. 1 (January 1971): 90–96.

Chapter 49

Radio-Telescope Antennas

William A. Imbriale
Dayton L. Jones

*Jet Propulsion Laboratory
California Institute of Technology*

CONTENTS

49.1	RADIO-TELESCOPE INTRODUCTION	49-2
49.2	SINGLE-APERTURE RADIO TELESCOPES	49-5
49.3	THE TWO-ELEMENT INTERFEROMETER	49-10
49.4	APERTURE SYNTHESIS	49-15
49.5	MAJOR RADIO TELESCOPES	49-21
49.6	CONCLUSIONS	49-24

49.1 RADIO-TELESCOPE INTRODUCTION

In contrast to optical telescopes, which produce visible light images, a radio telescope detects radio waves emitted by radio sources, typically by means of a large reflector antenna, or arrays of them. The first extraterrestrial radio source was detected by Karl Jansky, an engineer with the Bell Telephone Laboratories, and announced in *The New York Times* on May 5, 1933:

NEW RADIO WAVES TRACED TO CENTRE OF THE MILKY WAY. Mysterious static reported by K. G. Jansky, held to differ from cosmic ray. Direction is unchanging. Recorded and tested for more than a year to identify it as from Earth's galaxy. Its intensity is low. Only delicate receivable able to register—no evidence of interstellar signaling.

Today, we view this announcement of Jansky's paper, entitled "Electrical Disturbances Apparently of Extraterrestrial Origin," presented at the meeting of the U.S. Committee of the International Union of Radio Science (URSI) in Washington, D.C., as the beginning of radio astronomy. In Jansky's honor, astronomers named the unit of radio flux the jansky (which equals $10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$).

Grote Reber was one of the pioneers of radio astronomy. Using a 9-m telescope constructed in 1937, he was instrumental in repeating Karl Jansky's pioneering but somewhat simple work, and conducted the first sky survey in the radio frequencies. After World War II, substantial improvements in radio astronomy technology were made by astronomers in Europe, Australia, and the United States, and the field of radio astronomy began to blossom.

One of the most notable radio astronomy developments came in 1946, with the introduction of radio interferometry (see, for example, "Solar Radiation on 175Mc/s"¹ by Martin Ryle's group at Cambridge University, which was awarded a Nobel Prize for this, and later aperture synthesis work). Another major development was the Lloyd's mirror interferometer, developed independently in 1946 by Joseph Pawsey's group at the University of Sydney.²

In the early 1950s, the Cambridge Interferometer mapped the radio sky to produce the famous *Second Cambridge Catalogue of Radio Sources (2C)* and *Third Cambridge Catalogue of Radio Sources (3C)*³ surveys of radio sources. In the late 1950s and early 1960s, the largest single-dish radio telescope in the world was the 76-m telescope at Jodrell Bank, England, which became operational in 1957. It was one of many radio telescopes constructed during the middle of the 20th century, and has been surpassed by more modern telescopes and arrays of telescopes.

This chapter describes the basic types and functions of modern radio telescopes and highlights several of the more notable instruments.

Position and Coordinates

The accurate position of a radio source is necessary to distinguish the source from others and to assist in its identification with optical objects when possible. The position is conveniently expressed in *celestial equatorial coordinates*: *right ascension*, α ; and *declination*, δ . The poles of this coordinate system occur at the two points where Earth's axis, extended, intersects the celestial sphere. Midway between these poles is the *celestial equator*, coinciding with Earth's equator, expanded.

The declination of an object is expressed in degrees and is the angle included between the object and the celestial equator. It is designated as a positive angle if the object is north of the equator and negative if south. For example, at the Earth's equator, a point directly overhead (the zenith) has a declination of 0 deg, while at a north latitude of 40 deg, the declination of the zenith is +40 deg.

The *meridian* is a great circle passing through the poles and a point directly overhead (the zenith). The *hour circle* of an object is the great circle passing through the object and the poles. The *hour angle* of the object then is the arc of the celestial equator included between the meridian and the object's hour circle. This angle is usually measured in hours.

A reference point has been chosen on the celestial equator. It is called the *vernal equinox*. The arc of the celestial equator included between the vernal equinox and the object's hour circle is termed the *right ascension* of the object. It is measured eastward from the vernal equinox and is usually expressed in hours, minutes, and seconds.

The right ascension and declination of an object define its position in the sky, independent of the Earth's diurnal rotation. However, because of the slow wobble (precession) in the direction of Earth's rotation axis, there is a gradual change in these coordinates for a fixed object in the sky, the change completing one cycle in 26,000 years. Thus, the right ascension and declination of an object will again be the same as they are now in 26,000 years. To be explicit, it is necessary to specify the date to which the right ascension and declination refer. This date is called the *epoch*. At present, the *epoch* 2000.0 is commonly used in catalogs of source positions (that is, the right ascension and declination are those of January 1, 2000, which are then precessed to the current date).

Polarization and the Stokes Parameter

Any monochromatic wave is elliptically polarized, i.e., the end point of its electric vector at each point in space periodically traces out an ellipse or one of its special forms, viz. a circle or straight line. On the other hand, a polychromatic wave can be in any state of polarization, ranging from the elliptically polarized state to the unpolarized state, where the end point of the electric vector moves quite irregularly. Cosmic radio waves are generally in neither of these two extreme states but rather in an intermediate state containing both elliptically polarized and unpolarized parts. Such a wave is said to be "partially polarized" and is described by four parameters first introduced by Sir George Gabriel Stokes in 1852.^{4,5}

Given a specific coordinate system, it is possible to express the Stokes parameters in terms of the orthogonal electric field components of the incident radiation. Considering the simplest case of a monochromatic completely polarized wave, the orthogonal components of the incident radiation can be written as

$$\begin{aligned} E_x &= E_1 \sin(\omega t - \delta) \\ E_y &= E_2 \sin(\omega t) \end{aligned} \quad (49-1)$$

where δ is the phase difference between E_x and E_y , and E_1 and E_2 are real amplitudes. The four Stokes parameters are

$$\begin{aligned} I &= E_1^2 + E_2^2 \\ Q &= E_1^2 - E_2^2 \\ U &= 2E_1E_2 \cos(\delta) \\ V &= 2E_1E_2 \sin(\delta) \end{aligned} \quad (49-2)$$

A telescope measures radiation over a finite bandwidth so that the net polarization is the result of averaging many monochromatic components with independent

polarization properties. In that case, we can only measure average polarization properties. Thus, for partially polarized waves, the quantities E_1 , E_2 , and δ are time dependent, so time averages must be taken to find the Stokes parameters:

$$\begin{aligned} I &= \langle E_1^2 \rangle + \langle E_2^2 \rangle \\ Q &= \langle E_1^2 \rangle - \langle E_2^2 \rangle \\ U &= 2 \langle E_1 E_2 \cos(\delta) \rangle \\ V &= 2 \langle E_1 E_2 \sin(\delta) \rangle \end{aligned} \quad (49-3)$$

The I parameter is a measure of the power of the polarized radiation, Q and U are measures of linear polarization, and V is a measure of circular polarization. For a completely unpolarized wave, $Q = U = V = 0$.

Since the trigonometric functions above can never be greater than 1, and since waves can add in and out of phase (interfere constructively or destructively), we see that

$$I^2 \geq Q^2 + U^2 + V^2 \quad (49-4)$$

We now define the degree of polarization as

$$H = \frac{\sqrt{Q^2 + U^2 + V^2}}{I} \quad (49-5)$$

For the monochromatic wave $I^2 = Q^2 + U^2 + V^2$, the four parameters are not independent.

For fully polarized waves, the degree of polarization is 1, and for unpolarized waves the degree of polarization is 0.

Stokes parameters are particularly useful for describing the cosmic microwave background radiation (CMBR) as it is largely unpolarized, with a small polarized anisotropy on the order of one part in 10^6 . Thus, the partial polarization of the CMBR is expressed in terms of the Stokes parameters (I , Q , U , and V), which constitute the observables for most CMBR polarization experiments. Note that the CMBR is not expected to have circularly polarized anisotropies, so the measure of V can actually be used as a null test in polarimeter calibration.

Stokes parameters have units of power although CMBR experiments measure the Stokes parameters as a temperature at points on the sky (the two units are related by a constant conversion factor). Thus, polarization experiments usually refer to the temperature Stokes parameters Q_T , U_T , I_T , and V_T . The Q and U parameters are separated by 45 deg, and a 90-deg rotation of either parameter results in its inverse (so Q rotated by 90 deg gives $-Q$). It is possible to redefine the linear polarization Q and I Stokes parameters as

$$\begin{aligned} I &= \frac{T_x + T_y}{2} \\ Q &= \frac{T_x - T_y}{2} \end{aligned} \quad (49-6)$$

where T_x and T_y are the temperature measurements of the x and y orthogonal polarization components. Note that U can be derived by rotating the detection axes by 45 deg. There are a number of experiments under way to measure the polarization properties of the CMBR (see, for example, Leitch et al and Piacentini et al^{6,7}).

49.2 SINGLE-APERTURE RADIO TELESCOPES

The basic properties of single-aperture radio telescopes are described here, including common antenna geometries and feed systems.

Beamwidth and Sidelobes

The beam of a single-aperture telescope is the Fourier transform of the aperture illumination. Consequently, one of the major requirements in the design of telescope feeds and geometries is to control the illumination of the primary reflecting surface to produce a beam with the desired characteristics. There is a fundamental conflict between maximum sensitivity along the main beam axis and the level of sidelobes, because the Fourier transform of a function with sharp edges will contain power over a wide range of spatial frequencies.

For example, maximum telescope sensitivity occurs when the power received by each unit of collecting area is added equally, i.e., none of the power falling on the telescope aperture is wasted. This is equivalent to saying that the aperture is uniformly illuminated by the feed. In one dimension, uniform illumination is a rectangular function that is zero outside of a width equal to the diameter of the telescope's primary reflector. But the Fourier transform of a rectangular function is a sinc function whose central peak has a width inversely proportional to the width of the rectangle.⁸ However, a sinc function also has significant power outside of the main peak in the form of gradually decreasing oscillations (ringing). Thus, a telescope with uniform illumination has maximum sensitivity along the main beam (the central peak of the sinc function) but also has large sidelobes (the multiple maxima of the sinc function on either side of the central peak) that respond to radiation from other directions.

For applications where main-beam sensitivity is the primary goal, such as observations of point sources (spacecraft or natural radio sources whose angular size is much smaller than the telescope's beamwidth), uniform illumination is the proper goal. An aperture illumination pattern with perfectly sharp edges is not possible, but with careful design of the feed and reflector shapes, a good approximation to this can be achieved.

In many radio astronomy observations, the radio emission comes from a region that is not angularly small compared to the width of the telescope beam. In this case, large sidelobes can cause serious confusion, since the telescope is responding simultaneously to multiple areas of the radio source. For observations where this is a concern, such as mapping of extended emission regions or observing a weak radio source with stronger sources nearby in the sky, it is worth trading some sensitivity for much lower sidelobes.

As an extreme example of this, consider a telescope illuminated using a Gaussian function. The Fourier transform of a Gaussian is just another Gaussian whose width is inversely proportional to the width of the first Gaussian. Here we have a telescope beam whose amplitude decreases smoothly to a very low level with no ringing. Of course, true Gaussian illumination is not possible because the function will be cut off at some point by the edge of the telescope aperture. On the other hand, the Fourier transform of a more realistic truncated Gaussian contains some high spatial-frequency components, but if the truncation occurs far enough from the central peak so that the Gaussian amplitude is already very small, the resulting telescope beam will be a reasonable approximation of a Gaussian. In this case, however, much of the telescope collecting area is being used very inefficiently; the power received by the surface near the edge of the primary reflector is largely rejected by the feed.

Compromises between the extreme cases of aperture illumination described above have been implemented. In many cases, a cosine-squared illumination function has been adopted because it provides higher main-beam sensitivity than Gaussian illumination and lower sidelobes than uniform illumination.

In general, illumination functions that drop quickly and smoothly to zero at the edge of the primary reflector are desirable not only because they reduce sidelobes but also because they do not see radiation coming from beyond the edge of the dish (thermal emission from the ground, for example). Illumination patterns that do not drop to zero at (or inside) the reflector edge are said to spill over the reflector, and will respond to radiation from the ground. This, in turn, increases the background noise level of the observations.

Directivity, Gain, and Effective Aperture

The directivity and gain of a radio telescope are both measures of the ratio between the telescope's response to radiation from the direction of the main beam to the response to radiation from other directions. This ratio depends on the width of the telescope's main beam, which in turn is inversely proportional to the width of the telescope aperture in units of the observing wavelength. The constant of proportionality depends on the aperture illumination function but is normally close to 1.

If the telescope beam is described as a normalized power pattern in spherical coordinates, then the beam (or pattern) solid angle in steradians is the integral of the power pattern over a sphere (4π steradians). If sidelobes in the power pattern are negligible, the beam solid angle is equal to the main beam solid angle. In many cases, the beam solid angle will be close to the solid angle of the main beam within the half-power points. Directivity is equal to $4\pi/\text{beam solid angle}$. Thus, the directivity of a telescope is proportional to its diameter squared. For an isotropic antenna, directivity = 1 for all directions. (This corresponds to the Fourier transform of a delta function.) The gain of a telescope is the directivity converted to logarithmic units (usually dB with respect to an isotropic antenna, or dBi).

The half-power beamwidth of a telescope whose reflector diameter is N wavelengths will be approximately $1/N$ radians (in one dimension). This corresponds to a beam solid angle of $\pi/(4N^2)$ steradians in two dimensions. Since the beam solid angle is also equal to $4\pi/\text{directivity}$, we have directivity = $16N^2 = 16 (\text{diameter/wavelength})^2 = 4\pi (\text{area})/(\text{wavelength})^2$. Here, the area is the effective area of the telescope.

The effective area is equal to the geometric area in the case of perfect uniform illumination and no losses, but in reality, the effective area will be smaller than the physical area. The ratio of effective area to physical areas is the aperture efficiency. Typical radio telescopes have aperture efficiencies of 0.5–0.7, but lower values can occur near the low- or high-frequency ends of their designed frequency range.

Common Antenna Geometries

The vast majority of telescopes used for radio astronomy are axially symmetric reflectors. A number of other designs such as cylindrical parabolic reflectors, fixed spherical and parabolic reflectors, and large arrays of low-gain antennas are described in Section 49.5.

The main differences between symmetric reflecting telescopes are the number of reflecting surfaces and the location of the focal point(s). The simplest geometry is a single parabolic reflector with a feed located at the focal point. This is commonly referred to as a prime-focus telescope. It is used chiefly at low frequencies, where the long wavelengths require a physically large reflector to avoid significant diffraction effects. In this regime, any additional reflecting surfaces would also need to be physically large.

At higher frequencies, either the dual-reflector Cassegrain or Gregorian geometry can be used. In both cases, a secondary reflector (also called a subreflector) is used in the optical path between the feed and the primary parabolic reflector. The main difference in the two geometries is the shape of the subreflector: In Cassegrain geometry, the subreflector is a hyperbola with the feed at one focal point of the hyperbola (the secondary focus) and the other focal point of the hyperbola at the primary parabola focal point. In Gregorian geometry, the

subreflector is a concave ellipse with the feed at one focal point and the other focal point at the ellipse focal point. Gregorian geometry allows a low-frequency feed to be placed at the first focus and operated in prime-focus mode without removing the subreflector. Cassegrain geometry requires that the subreflector be removed for a prime-focus feed to be used.

It is possible to improve the efficiency of such systems by changing the shape of the subreflector to optimize the primary aperture illumination (for example, to make it more uniform for maximum telescope gain). If this is done, the primary reflector shape must also be changed to compensate for the additional phase changes caused by the new subreflector shape and to restore uniform phases across wavefronts. Telescopes that have been optimized in this way are called shaped antennas and can achieve aperture efficiencies of about 0.8.

A variation on the two-reflector geometries is a nonsymmetrical, offset section of a parabola for the primary reflector shape. This allows the subreflector (and the prime focus) to be located out of the beam path and thus avoids blockage and scattering of incoming radiation. Figure 49-1 shows one of the Allen Telescope Array antennas (under construction in Northern California) before the feed and ground shield are mounted. The feed will go near the bottom of the main reflector, pointing towards the subreflector. The ground shield will be below the feed, to prevent sidelobes from seeing radiation from below.

Some telescopes use multiple reflectors to steer the beam from the subreflector to a location where the feeds and receivers can remain stationary and can be changed rapidly. These beam-waveguide (BWG) geometries are particularly appropriate when high power transmission is needed, since the heavy transmitter does not need to be mounted on the moving telescope structure. An example of the BWG design used by NASA's Deep Space Network is shown in Figure 49-2.



FIGURE 49-1 The ATA offset Gregorian optics before the addition of the feed and ground shield

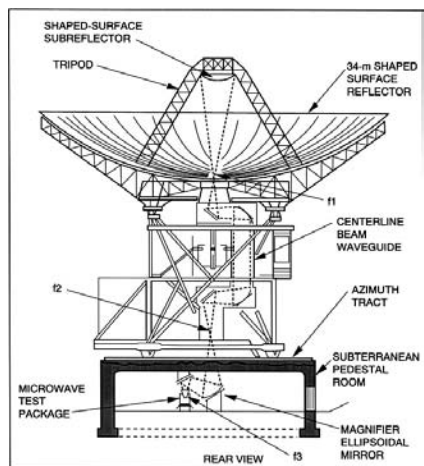


FIGURE 49-2 Deep Space Network research and development antenna

Radio telescopes designed for very high frequencies (millimeter and submillimeter wavelengths) often use additional reflecting surfaces to allow receivers to be located along the elevation axis and thus remain fixed as the telescope moves in elevation. At very high frequencies, the subreflector can be physically small and, consequently, it is practical to move it rapidly to switch the telescope beam between on-source and off-source positions. This allows better calibration by canceling slowly varying atmospheric effects.

Common Feed Types

There are four broad classes of feeds: (1) narrowband, (2) multiband, (3) wide-band, and (4) phased-array. Traditionally most telescopes for radio astronomy have been equipped with narrowband conical horn feeds. These feeds can provide very high performance but at the expense of flexibility. The frequency range covered by horn feeds is usually no more than a factor of 2. Horn feeds are usually designed with ridges or corrugations along the inside of them to control the profile of the radiation pattern. They can have high gain (narrow beamwidths), which makes them a natural match to high-frequency telescopes with small subreflectors. They can also be designed to have low gain (for prime focus) by making the output matching section a very-wide-angle cone. However, at low frequencies, all horn feeds tend to be physically large, which makes them unsuitable for cryogenic cooling. At high frequencies, horn feeds can be placed inside Dewars and cooled to reduce noise associated with ohmic losses in the feed material. For high-gain feeds, the vacuum window in the Dewar needs to be only slightly larger than the diameter of the feed. An alternative type of narrowband feed for low frequencies is a pair of crossed dipoles or a set of four dipoles in a square configuration, mounted over a ground plane (and sometimes also surrounded by a short, cylindrical cavity wall). These feeds are appropriate for prime-focus mounting, are smaller than a horn feed for the same frequency, but have higher sidelobes and consequently lower performance.

A special case of a narrowband feed is a line feed, which can be used to compensate for spherical aberration caused by using a spherical primary reflector. Line feeds are designed so that the propagation of velocity of signals along the feed is equal to the speed of signal convergence along the linear focus of a spherical reflector. Line feeds are intrinsically very narrowband and are insensitive directly along their axis.

Multiband feeds are usually combinations of two or three narrowband conical feeds that are mounted coaxially. An example of this type of feed is shown in Figure 49-3.

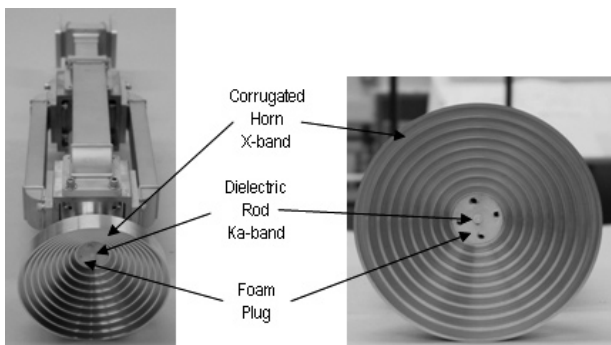


FIGURE 49-3 Multiband feed

Wide-band feeds sacrifice some performance in exchange for exceptional flexibility. Most wide-band feed designs are based on log-periodic structures of some sort, in which radiation at a given frequency couples with the part of the structure that is resonant at that frequency and then travels along the structure as a balanced feedline to the amplifier input. Log-periodic feeds can cover frequency ranges exceeding 20:1, with a nearly constant beamwidth throughout the whole range. However, the beamwidth is normally very large (60–90-deg full opening angle), and sidelobe levels are relatively high. Log-periodic feeds can be made to have higher gain if they have a small opening angle (i.e., are tall and slender). In this case, several elements are nearly resonant at the same frequency and can interact like the elements of a Yagi antenna. However, the phase center in this type of feed moves along the structure with frequency, so some of the wide-band advantage is lost because only part of the feed structure can be near the telescope focal point at a given time. Newer designs avoid this problem by using a very large opening angle for the log-periodic antenna (short and wide structure). The change in phase center with frequency is kept small because the entire structure is very short. In this case, very low feed gain is unavoidable.

Phased-array feeds are composed of many antenna elements whose signals are combined with appropriate phase shifts to create a desired beam profile. Because the signal from each element can be combined with other signals an arbitrary number of times, phased arrays can create multiple beams simultaneously and thus cover a much larger area of sky than a single feed can. The ability to control phases can also be used to correct for wavefront distortions caused by errors in the telescope's reflecting surfaces. The main limitations of phased array feeds are that they require separate low-noise amplifiers for each antenna element, thus increasing the cost. And they are difficult to use over wide frequency ranges because the physical spacing between elements is optimal for only a narrow range of frequencies.

Most feeds are inherently linearly polarized, while most radio astronomy observations use circular polarization to eliminate the effects of Faraday rotation in the ionosphere. Circular polarization can be obtained from a pair of orthogonal linear feeds by combining the signals from them with a 90-deg phase difference. With this approach, the useful frequency range may be limited by the polarizer more than by the feed. Some feeds, such as helical feeds, are inherently circularly polarized.

Calibration of Telescope Sensitivity

For radio astronomy, the important parameter for telescope sensitivity is system equivalent flux density (SEFD), which can be calculated by observing a radio source of known flux density and measuring the increase in system temperature due to the source. The ratio of the temperature increase over the source flux density is the telescope sensitivity in K/Jy , which depends only on telescope size and efficiency.

The increase in system temperature caused by a radio source is the antenna temperature of the source. This can be measured in kelvins if the system temperature is calibrated using hot and cold loads in front of the first amplifier. If the off-source system temperature is T_{sys} , the antenna temperature T_{ant} is

$$T_{ant} = (P_{on} - P_{off} / P_{off}) \times T_{sys} \quad (49-7)$$

where P_{on} and P_{off} are the receiver output power level with the telescope on source and off source. Having the antenna temperature in kelvins for a radio source of known flux density, we can determine the aperture efficiency of the telescope.

To include the effect of system temperature on sensitivity, we can use the K/Jy value to convert the system temperature T_{sys} into flux density units (Jy). This is the SEFD, and it combines

the effects of telescope size, efficiency, and receiver performance (system temperature). It is a useful way to compare the sensitivities of different telescopes; smaller values of SEFD are better because they imply higher sensitivity (a combination of high K/Jy and low T_{sys}).

49.3 THE TWO-ELEMENT INTERFEROMETER

The basic characteristics of the two-element interferometer are described here. The two-element interferometer provides the fundamental building block for astronomical observations.

Electric Field Correlations

An interferometer is an instrument that measures the correlation of electric fields at spatially separate locations. These correlations contain information about the sky brightness distribution on angular scales that are inversely related to the spatial scale over which the field correlation between points is measured. Thus, an interferometer that measures the cross-correlation of electric field received at points separated by a small physical distance is sensitive to emissions from the sky on large angular scales, and an interferometer that measures cross-correlations between points physically far apart is sensitive to small angular-scale structures in radiation from the sky.

The cross-correlation of two electric fields, E_1 and E_2 , is $\langle E_1 E_2 \rangle$ so it is equivalent to multiplying the received signals from one telescope by the received signals from a second. However, the cross-correlation will be significant only if it is the product of samples from different parts of the same wavefront. That means that any propagation delay between the time a given wavefront reaches one telescope and the time it reaches the second telescope must be compensated for before the signals from the two receivers are multiplied by each other.

Let T be the geometric delay between the two telescopes (equal to the dot product of the baseline vector between the two telescopes and a unit vector in the direction of the radio source being observed), and let H be the Hilbert transform (corresponding to a 90-deg phase shift). Then the cross-correlation between voltages E_1 and E_2 from telescopes 1 and 2 is

$$C = \langle E_1(t) E_2(t + T) \rangle + i \langle H[E_1(t)] E_2(t + T) \rangle \quad (49-8)$$

Note that the cross-correlation defined here is a complex function. In practice, it is obtained by delaying the signal from telescope 1 by the geometric delay, T ; multiplying (and averaging) the delayed E_1 with E_2 ; and simultaneously multiplying (and averaging) E_1 shifted by 90 deg with E_2 . The circuit that performs these tasks is called a complex correlator, and it is a basic component of all modern interferometers. The complex sum of the two independent outputs of a complex correlator is the visibility function, which differs from the complex cross-correlation function only in details of amplitude calibration.

The Visibility Function

The visibility function is the Fourier transform of the sky brightness distribution. As mentioned above, it is a complex function. Because sky brightness is a positive real function, its Fourier transform must be a Hermitian function: (1) the visibility function is complex, (2) the real part is symmetric about the origin, and (3) the imaginary part is antisymmetric about the origin. The visibility measured by an interferometer is a function of sky brightness distribution and the projected length and orientation of the interferometer baseline. It is *not* a function of the absolute positions of the telescopes but only of the

projection of the baseline vector between the telescopes in the direction of the source. Each observation with an interferometer provides one sample of the visibility function. From a number of such measurements, the strength and position of radio sources can be deduced.

At this point, it is useful to recall a few mathematical properties of Fourier transforms. These include linearity (the visibilities of multiple radio components add), shift (changing the position of a source on the sky produces a phase gradient in the visibility function), and similarity (larger radio sources have more compact visibility functions).

It is traditional, and also helpful, to imagine that measurements of the visibility function are all made on a plane. Then we can define a coordinate system that has one axis orthogonal to this plane. The coordinates of this system are (u, v, w) , with w orthogonal to the plane in which the visibility measurements are made. In other words, w points toward the radio source being observed. By convention, u points east, and v points toward the north celestial pole.

$$V_v(u, v) = \iint \frac{I_v(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2i\pi(ul + vm)} dl dm \quad (49-9)$$

The sampling of the visibility function by interferometers can be quantified by specifying the position in the projected baseline vector in the (u, v) plane. The more thoroughly the (u, v) plane is sampled, the more completely the visibility function is known, and the more accurate our knowledge of the sky brightness distribution becomes. Recall that the amplitude of the visibility function tells us about the radio source structure, and the phase tells us about the source position on the sky.

Geometric Delay

The geometric delay of an interferometer depends only on the length and orientation of the baseline with respect to the direction of the radio source being observed. If \mathbf{B} is the baseline vector between two telescopes, and \mathbf{S} is a unit vector in the direction of the source, the geometric delay is just the dot product $(\mathbf{B} \cdot \mathbf{S})/c$. As the interferometer tracks a source across the sky, $\mathbf{B} \cdot \mathbf{S}$ changes and, consequently, the geometric delay changes. For a perfect system, we can introduce an exact delay correction to keep the fringe phase constant as a function of time.

In real instruments, there is always some error in our knowledge of the baseline vector, and often there is also some uncertainty in the position of the radio source. These errors contribute to an uncertainty in the calculated geometric delay of

$$\Delta\tau = \sqrt{(\sin\theta/c)^2 \Delta B^2 + (B \cos\theta/c)^2 \Delta\theta^2} \quad (49-10)$$

If the uncertainty in delay is large compared to the Nyquist sampling⁹ interval in delay (the inverse of twice the observing bandwidth), then a search over multiple delays is needed to determine the delay that produces the highest visibility amplitude. This is called fringe searching and is most often needed when the separation between telescopes is very large (hundreds to thousands of kilometers).

If we know the sky positions of several radio sources very well, we can use this knowledge to determine the baseline vector of an interferometer. Let ΔB be the error in our assumed baseline vector \mathbf{B} (in wavelengths), let δ be the declination of a radio source, and let H be the hour angle of the radio source (the distance in right ascension from transit, where $H = 0$). The interferometer phase as a function of hour angle is $\phi(H) = 2\pi(\mathbf{B} \cdot \mathbf{S})$ for a perfect system. The baseline error $\Delta B = \Delta B_x, \Delta B_y, \Delta B_z$ causes a change in the baseline phase of

$$\Delta\phi(H) = 2\pi \Delta B_x \cos(\delta) \cos(H) - 2\pi \Delta B_y \cos(\delta) \sin(H) + 2\pi \Delta B_z \sin(\delta) \quad (49-11)$$

We are ignoring any constant instrumental phase. By measuring $\Delta\phi(H)$ for sources at multiple declinations, we can solve for ΔB_x , ΔB_y , and ΔB_z (and a constant instrumental phase as well).

Fringe Rotation

Radio interferometers normally convert the received high radio frequencies to a lower, intermediate frequency for transmission to the correlator. As a result, when the geometric delay compensation is applied as part of the cross-correlation process, it is applied at a different frequency than the frequency the geometric delay occurred at. The result is a phase error, which can be corrected by adjusting the phase of the local oscillator at the telescope whose signal undergoes the geometric delay compensation.

Phase Switching

Systematic errors at very low fringe rates can limit the accuracy of interferometer measurements. These errors include interference picked up at the intermediate or baseband frequencies, offsets in digitizer thresholds, or cross-talk between correlators. By applying 180-deg phase shifts using orthogonal switching waveforms to the signals from the telescopes, we can suppress this type of error. An example of orthogonal digital waveforms are the Walsh functions, which are used for phase switching of the local oscillators at the Very Large Array (New Mexico) telescopes.^{10,11} When integrated over many phase switching cycles, signals entering the system before the first mixer (where the local oscillator signal is applied) are unaffected by phase switching, while signals that enter the system after the first mixer are suppressed.

Digital Cross-Correlation

Nearly all modern correlators are digital, but the process of digitizing an analog signal causes loss of information. For wide-bandwidth signals, it is often impractical to use many bits per sample when digitizing and, consequently, it is necessary to correct the measured visibility amplitudes for the loss in signal-to-noise ratio (SNR) caused by the sampling. For 1-bit sampling, as used for interferometers in which the widest possible bandwidth is desired, the sampling loss is a factor of about 1/3. Even for 2-bit (4-level) sampling, the loss is greater than 10 percent.

It is also necessary to make a digital correction to both the visibility amplitude and phase, which depends on the signal-to-noise ratio (SNR) and the number of bits used to digitize the signal. Van Vleck¹² first noted that digitization causes an increase in correlation at low correlation levels and a predictable nonlinearity at high correlation levels. The correction is applied separately to the real and imaginary output of a complex correlator, so the correction affects the visibility phase as well as the amplitude.

Calibration

Possibly significant error sources, in the order they occur along the signal chain, are

- **Ionospheric Faraday Rotation** The linear polarization angle of radiation passing through the ionosphere is rotated by a difference in propagation velocity between left and right circular polarization (birefringence). This effect is proportional to wavelength

squared and, consequently, can be measured and corrected by using widely spaced frequencies to measure the frequency dependence of the polarization angle. It is also possible to determine ionosphere electron density from Global Positioning System measurements.

- **Troposphere Effects** Variations in water vapor in the troposphere cause changes in amplitude due to opacity, and changes in phase due to excess path length. These changes can be a dominant error source at high frequencies. Water vapor radiometers, which measure the strength of water line emission, can help to calibrate these changes, but the best solution is usually antenna-based self-calibration (discussed in Section 49.4).
- **Parallactic Angle** This is a time-varying angle between the feed polarization response and the sky; it is important for polarization measurements. It can be calibrated by observing radio sources with known linear polarization position angles. The parallactic angle is defined as

$$\chi(t) = \arctan \left(\frac{\cos(l) \sin(h(t))}{\sin(l) \cos(\delta) - \cos(l) \sin(\delta) \cos(h(t))} \right) \quad (49-12)$$

l = latitude, $h(t)$ = hour angle, δ = declination

- **Antenna Voltage Pattern** The antenna voltage pattern is important only for wide-field observations. It can be determined from direct measurement or from modeling of the feed and mechanical structure of the telescope.
- **Electronic Gain** This includes a large number of amplitude and phase effects introduced by imperfect electronics (in amplifiers, mixers, filters, digitizers, etc.) but excluding frequency-dependent effects. Calibration usually requires a stable broadband noise source that can be added to the signal path before the first active device. Measurement of the total system temperature combined with the telescope sensitivity in units of K/Jy allows the amplitude of interferometer fringes (i.e., the amplitude of the visibility function) to be determined.
- **Bandpass Response** Both amplitude and phase response can vary across the observed bandwidth due to instrumental or—at millimeter and submillimeter wavelengths—atmospheric effects. The bandpass response can be calibrated by observing a strong continuum radio source.
- **Geometric Compensation** Errors in the baseline vector can be corrected using observations of radio sources having well-known positions. If the calculated geometry is imperfect, there will be a residual baseline phase that varies with the angle between the source direction and the interferometer meridian. After the residual phases of sources are measured as they move across the sky, the resulting data can be used to solve for corrections to the baseline vector via a least-squares fit.

Sensitivity

The root-mean-square (rms) noise level in the output of an interferometer (in janskys) is given by

$$\Delta S = \frac{1}{\eta_s} \times \frac{\text{SEFD}}{\sqrt{2 \times \Delta \nu \times \tau_{\text{acc}}}} \quad (49-13)$$

where SEFD is the system equivalent flux density of the interferometer, η is the efficiency, $\Delta \nu$ is the observing bandwidth, and τ is the integration time.

Model Fitting

A single baseline interferometer cannot produce sufficient sampling of the visibility function to result in a true radio image of the sky. However, it is still possible to determine the basic characteristics of a radio source with model fitting. This technique involves relatively simple geometric modeling for one- or two-point sources, Gaussian or uniform disks, edge-darkened or edge-brightened disks, as examples.

The model parameters (flux density, position, diameter, axial ratio for elliptical sources, etc.) are adjusted to minimize the difference between the Fourier transform of the model and the observed visibility data.

With accurately calibrated fringe amplitudes, it is possible to determine source flux densities and sizes (and sometimes shapes) with model fitting. The orientation and angular separation of double sources can also be found and changes monitored in this way. With the addition of fringe-phase data, the sky positions of sources can be found with model fitting (although it is much more difficult to accurately calibrate interferometer phase data than amplitude data).

Astrometry

Positional information is contained in the phase part of the visibility function. The residual phase on an interferometer baseline tells us how far the true signal delay between the two telescopes differs from the calculated delay. If the interferometer baseline is well known, this in turn tells us how far the radio source is from its assumed position. Because baseline phases can be difficult to calibrate and use directly, it is sometimes advantageous to use instead the phase differences between different frequencies. This observable is immune to ambiguities in total phase values.¹³ The derivative of phase with frequency is the group delay, and the accuracy with which it can be measured depends on the total range of frequencies used. If the total bandwidth that can be observed is too small for a good group delay measurement, it can be split between two narrower bands that are spaced farther apart in frequency. This is called bandwidth synthesis, and is used for position measurements on long baselines, where it is very hard to avoid ambiguities in the total number of turns of baseline phase.

Baseline phases can be used directly to measure angular separations between sources. These phase-referenced measurements do not provide absolute source positions but can provide very precise relative positions. In phase referencing, one source is used to determine instrumental and atmospheric effects, and the phase difference of the second source gives us the position difference. This approach works well when the sources are angularly close enough that the two lines of sight from each telescope pass through the same regions of the atmosphere and ionosphere and if at least one of the sources is strong enough that the integration time needed to obtain a high SNR is shorter than the time scale of atmospheric propagation changes. All other instrumental errors cancel in the differencing.

If, however, the second source is too weak to detect in a short integration time, the measured phases from the stronger source can be used to extend the coherent integration time on the weak source. In this mode of phase referencing, the coherent integration time can be extended to hours instead of the usual seconds to minutes.

The limitation to this technique is the lack of precision with which the phases on the strong source can be interpolated to the times when the weaker source is being observed.

The precision of astrometric position measurements is the angular resolution of the interferometer (the observing wavelength divided by the baseline length) divided by twice the SNR. It is possible to obtain high precision with long-baseline interferometers and moderate SNR or with short baseline interferometers and high SNR. Ideally, of course, one has both. Current intercontinental interferometers can measure relative positions with a precision better than 0.5 nanoradians for sources producing a signal only 10–20 times the thermal noise level.

49.4 APERTURE SYNTHESIS

The term *aperture synthesis* refers to the fact that a multibaseline interferometer can produce radio images with an angular resolution equal to that of the longest baseline. In other words, the interferometer can synthesize the resolution of an aperture with the same diameter as the array of telescopes. This section describes how aperture synthesis is used to form images.

(u, v) Plane and the Synthesized Beam

The (u, v) plane was introduced in Section 49.3. It is a plane orthogonal to the line of sight to a radio source in which the location of the projected baseline vector shows where the visibility function is being measured. Because the visibility function is Hermitian, a single baseline interferometer measures two points on opposite sides of the origin in the (u, v) plane.

A multitelescope interferometer with N telescopes has $N(N-1)/2$ simultaneous baselines (the factor of 2 is because the baseline between telescope 1 and 2 samples the same (u, v) points as the baseline between telescopes 2 and 1). The Fourier transform of the sampled points in the (u, v) plane, with each sampled point weighted equally, is the synthesized beam of the interferometer. This is the same as the point spread function in optical terminology. If N is large enough or the telescopes are moved to allow the baselines to sample the (u, v) plane reasonably well, the synthesized beam will have a well-defined peak along with a complex pattern of sidelobes. A synthesized beam with a single clear peak is necessary for aperture synthesis imaging.

Earth Rotation Aperture Synthesis

The position of the sampled points in the (u, v) plane moves as Earth rotates because the projected length and orientation of the baseline changes. The curve of sampled points in the (u, v) plane will be a section of an ellipse whose origin and axial ratio depend on both the baseline geometry and the position of the radio source.

The importance of Earth rotation aperture synthesis is that the constantly changing baselines of a multitelescope interferometer can, over a period of several hours, sample the visibility function (u, v plane) so well that an inverse Fourier transform of the visibility function can be calculated to produce an initial radio image of the region of sky being observed. The usual practice is to place the measured (u, v) points onto a uniform two-dimensional grid that extends at least a factor of 2 beyond the measured points to accommodate the fast Fourier transform (FFT) algorithm. This “gridding” of the measured data can be done by shifting, averaging, or interpolating, depending on the density of measurements and the spacing of points in the grid.

There are many options for weighting gridded visibility data. The most obvious is to weight every gridded data point equally. This provides a synthesized beam with maximum angular resolution, as it corresponds to uniform illumination in the single-aperture case. However, it also produces a beam with relatively high sidelobe levels (for the same reason), and lowers beam sensitivity somewhat because grid points in which several independent visibility measurements were averaged together have the same weight as grid points based on single measurements. The alternative is to weight each grid point by the number of measurements that it is based on—in effect, weighting all of the original measurements equally. This provides maximum sensitivity since no data are reduced in weight, but because most interferometers produce much denser sampling near the origin of the (u, v) plane than farther out in the long-baseline regions, the resulting synthesized beams are usually much broader than the angular resolution of the longest baseline. Weighting all visibility measurements equally does tend to

produce a synthesized beam with lower sidelobes and is often a better choice for the detection of extended radio emissions. If the goal of an observation is to detect or image very extended, low-surface brightness emissions, the synthesized beam can be further broadened by applying a Gaussian weighting function to the (u,v) data. This is called “tapering”; it effectively reduces the weight of data points far from the (u,v) origin. These points provide high angular resolution information, but they are insensitive to very extended emissions.

Of course, there will always be regions in the (u,v) plane that are not sampled and for which we will consequently have no observational constraints on the visibility function. These unsampled regions, if they are a significant fraction of the overall area of measurements in the (u,v) plane, can cause serious artifacts in the image produced by Fourier transforming the gridded visibility data. Historically, unsampled (u,v) points were set to zero before the Fourier transform was done. This was a computationally simple and often surprisingly effective solution, but it was certainly not an optimal way to approximate the visibility function at unmeasured grid points. More often, a convolving function is used to fill in missing values, but choosing this approach implicitly assumes that the unsampled regions are relatively small. Furthermore, in nearly all cases, the lack of dense and uniform measurement of the visibility function leads to image artifacts that need to be removed.

Instrument Calibration

Another cause of errors in radio images produced by aperture synthesis is imperfect calibration of the measured visibility function amplitude and phase. The correlator output can be calibrated to produce fringe or visibility amplitudes in janskys if the telescope sensitivities and signal processing losses are known. If C is the fraction of signal power from two antennas that is correlated (the correlation coefficient, or the fraction of bits correlated – $1/2$), then the calibrated fringe amplitude is given by

$$\begin{aligned}
 S_{\text{corr}} &= Cb \sqrt{(T_{\text{sys1}} T_{\text{sys2}}) / (T_{\text{ant1}} T_{\text{ant2}})} S_{\text{total}} \\
 &= Cb \sqrt{(\text{SEFD}_1 \text{ SEFD}_2)} \\
 &= Cb \sqrt{(T_{\text{sys1}} T_{\text{sys2}}) / (\text{DPJY}_1 \text{ DPJY}_2)} \\
 &= Cb 2k \sqrt{(T_{\text{sys1}} T_{\text{sys2}}) / (A_{\text{eff1}} A_{\text{eff2}})} \quad (49-14)
 \end{aligned}$$

where T_{sys} and T_{ant} are the system and antenna temperatures at a given telescope, S_{total} is the total flux density of the radio source in janskys, SEFD is the system equivalent flux density at a telescope, DPJY (degrees per jansky) is telescope sensitivity expressed in K/Jy , and A_{eff} is the effective area of a given antenna (the physical collecting area \times the aperture efficiency). Note that the antenna temperature is just the source flux density $\times K/\text{Jy}$ for the telescope ($K/\text{Jy} = A_{\text{eff}} / 2k$). The correction factor, b , accounts for system losses due to a number of causes, including nonrectangular bandpasses, mismatches in feed polarizations, phase fluctuations between local oscillators, signal digitization, and discrete steps in delay compensation; b is typically between 1 and 2 in well-designed interferometers and can be determined from observations of sources that are completely unresolved by the interferometer (for which $S_{\text{corr}} = S_{\text{total}}$).

The rms noise in a measurement of S_{corr} is given by

$$S_{\text{rms}} = b S_{\text{total}} \frac{1}{\sqrt{2BW \Delta\tau}} \sqrt{\frac{T_{\text{sys1}} T_{\text{sys2}}}{T_{\text{ant1}} T_{\text{ant2}}}} \quad (49-15)$$

where BW is the observing bandwidth in Hz, and $\Delta\tau$ is the coherent integration time in seconds. The value of $\Delta\tau$ is limited by either the atmospheric coherence time or the local oscillator coherence time. In terms of fractional frequency stability ($\Delta f/f$), the coherence time can be estimated $\Delta\tau = 1/[(\Delta f/f) \times \text{observing frequency in Hz}]$.

Phase calibration is more difficult than amplitude calibration because the baseline phase can change very rapidly due to propagation effects in the ionosphere (at low frequencies) or troposphere (at high frequencies). These phase fluctuations increase as baseline length increases, up to several kilometers, and saturate at longer baselines.

Self-Calibration

The most effective way to remove phase fluctuations from interferometer data is to remember that all of the major contributions to the phase errors are telescope based. That is, changes in the propagation along one telescope's line of sight—or local oscillator phase changes, or mispointing, or mechanical changes in the telescope structure, or receiver performance variations—all affect the signal from an individual telescope but have no effect on the signals from any other telescope. This means that errors in interferometer baseline data can be modeled as a product of errors associated with the individual telescopes^{14,15}:

$$V_{\text{obs}}(1,2) = V_{\text{err1}} \times V_{\text{err2}} \times V_{\text{src}}(1,2) \quad (49-16)$$

where V_{obs} is the observed complex visibility on the baseline between telescopes 1 and 2, V_{err} is a complex error associated with a given telescope, and $V_{\text{src}}(1,2)$ is the visibility due to the radio source that would have been measured in the absence of any telescope-based errors. (More correctly, we would use the complex conjugate of V_{err2} because the baseline phase is the difference of the telescope phases.) For a perfect interferometer, $V_{\text{obs}}(1,2) = V_{\text{src}}(1,2)$.

Now consider just the phase part of the visibilities and telescope errors:

$$\phi_{\text{obs}}(1,2) = \phi_{\text{err1}} - \phi_{\text{err2}} + \phi_{\text{src}}(1,2) \quad (49-17)$$

If we sum the observed baseline phases around a triangle of baselines between telescopes 1, 2, and 3, we get a quantity called the closure phase:

$$\begin{aligned} \phi_{\text{closure}}(1,2,3) &= \phi_{\text{obs}}(1,2) + \phi_{\text{obs}}(2,3) + \phi_{\text{obs}}(3,1) \\ &= [\phi_{\text{err1}} - \phi_{\text{err2}} + \phi_{\text{src}}(1,2)] + [\phi_{\text{err2}} - \phi_{\text{err3}} + \phi_{\text{src}}(2,3)] + [\phi_{\text{err3}} - \phi_{\text{err1}} + \phi_{\text{src}}(3,1)] \\ &= \phi_{\text{src}}(1,2) + \phi_{\text{src}}(2,3) + \phi_{\text{src}}(3,1) \end{aligned} \quad (49-18)$$

Note that the sum contains information about the radio source, but all the telescope-based error terms cancel. This means that the closure phase is a very robust observable, although it contains less total information than the individual baseline phases.

A similar trick can be used to produce a quantity that is independent of telescope-based amplitude errors:

$$A_{\text{closure}}(1,2,3,4) = \frac{|V_{\text{obs}}(1,2)| |V_{\text{obs}}(3,4)|}{|V_{\text{obs}}(1,3)| |V_{\text{obs}}(2,4)|} = \frac{|V_{\text{src}}(1,2)| |V_{\text{src}}(3,4)|}{|V_{\text{src}}(1,3)| |V_{\text{src}}(2,4)|} \quad (49-19)$$

Again, the telescope-based errors cancel in the ratio, and we are left with a closure amplitude that contains only information about the radio source.

For an array containing N radio telescopes, the number of independent closure phases and closure amplitudes is $\frac{1}{2}(N-1)(N-2)$ and $\frac{1}{2}N(N-3)$, respectively. These are both smaller than the number of baselines $\frac{1}{2}N(N-1)$. But it is clear that as N increases the amount of information contained in the closure, quantities becomes nearly equal to the information in the baseline amplitudes and phases. The only fundamental information loss associated with the closure phase and closure amplitudes is the absolute sky position and the absolute flux density of the radio source; the closure quantities are independent of these parameters.

The fact that major error sources in radio interferometers are telescope based and thus can be removed without destroying the information we want in the visibility function is the basis for self-calibration. For N telescopes, there are only $(N-1)$ complex errors to solve for (one telescope can be arbitrarily defined as a reference) and there are $\frac{1}{2}N(N-1)$ complex visibility measurements to constrain the solutions. Thus, we can use least-squares to adjust the telescope errors so that the resulting visibilities are as close as possible to the expected visibilities. But what should we expect the visibilities to be? Here we are helped by the fact that the sky is mostly empty at all but the lowest radio frequencies. Consequently, a model of the sky that consists of only a single source is often a good approximation. The Fourier transform of a point source is a visibility function with constant amplitude and zero phase everywhere. This is the usual starting point for self-calibration.

After a least-squares adjustment of the telescope errors, there will most likely be residual differences between the corrected visibilities and the assumed visibilities. This tells us that our model of the sky is imperfect. To achieve a better result, we can Fourier-transform the corrected visibilities to produce a more accurate model of the sky brightness, and then repeat the self-calibration process. Each iteration of self-calibration will produce both more accurate telescope calibration and a more accurate image of the sky. Eventually, the residual differences between the corrected and model visibilities will be so small and random that an image made from the residual visibility data will look like two-dimensional Gaussian noise with an rms amplitude close to the expected thermal noise limit. At that point, self-calibration has converged, and no further improvement in the telescope calibration of the radio image is possible.

Non-Closing Errors

We have seen that self-calibration of an interferometer array can remove telescope-based, or closing, errors. But are there errors that are really baseline dependent and that do not vanish in the closure quantities? There are, although in a well-designed interferometer nonclosing errors can be kept to a very low level. Nevertheless, they can limit the dynamic range of very high-quality radio images made with modern interferometers.

Examples of nonclosing errors are mismatched bandpass and polarization responses between telescopes. These produce additive errors instead of multiplicative errors, so the effect on all baselines to a given telescope will not be identical—it will depend on the telescopes at the other ends of the baselines. There are also subtle correlator errors that can be baseline dependent.

Nonclosing errors can be removed by observing a very strong radio source and measuring the bandpass response (for example) of each telescope accurately and then determining a bandpass calibration function for each telescope to produce a uniform response. Ideally, a very strong unresolved source can be used to calibrate the entire system at once because such a source should produce identical visibility amplitudes on all baselines. However, for interferometers with very long baselines, there are few, if any, radio sources that are completely unresolved.

Image Formation

The basic process of radio-image formation has been described already: grid, weight, and Fourier-transform the visibilities measured by an array of interferometers and, if necessary, use self-calibration to remove errors in the measured visibilities and repeat. The initial images produced by this process are a convolution of the actual sky brightness distribution with the synthesized beam of the array. Since the synthesized beam can be complex with high sidelobes, except for arrays with a very large number of telescopes, the resulting convolved image is usually not a good approximation of the sky brightness distribution. Indeed, the usual term for the synthesized beam is a “dirty beam,” which in turn produces a “dirty image” of the sky. This undesirable state of affairs can be remedied by removing the image artifacts created by the sidelobes of the dirty beam, as described below. Once a deconvolved image or model of the sky is produced, it can be convolved with a “clean beam”—the main peak of the synthesized beam without any of the sidelobes—to produce a cleaned image.

There is a fundamental relationship between the spacing of grid points in the (u,v) plane used for the Fourier transform and the angular separation between pixels in the image plane. To properly represent the radio source morphology, we need at least two pixels (and preferably three to five pixels) across the width of the main peak in the synthesized beam. This is another way of saying we need the image plane sampling to be finer than the angular resolution of the longest interferometer baselines. Otherwise, we are effectively wasting the long baselines in the array. Similarly, the image must contain enough pixels in each direction to cover the full extent of the radio emission. Otherwise, the large-scale structure information from the short baselines in the array will be wasted.

A given spacing in the image plane (ΔX) implies a grid spacing in the (u,v) plane of $1/(M * \Delta X)$, where M is the size of the grid along the axis in question. (The FFT need not use equal numbers of grid points in both dimensions, but this is the usual default.) Note that grid spacing in the (u,v) plane can be adjusted by changing the size of the FFT. There is a lower limit for M , based on the maximum size of the radio source being observed. In fact, $M * \Delta X$ must be at least twice the width of the radio emission to avoid aliasing in the FFT. Is there an upper limit? There is, because very fine sampling of the (u,v) plane means that a larger fraction of grid points is far from an actual measurement of the visibility function. Ideally, the distance between (u,v) grid points should be similar to the average spacing between visibility measurements.

The important point here is that there is a relationship between the range of angular structure that an interferometer array is sensitive to and the range of baseline lengths in the array. If there is a mismatch between the actual (u,v) sampling provided by the interferometer baselines and the range of angular resolutions needed, one side of the FFT or the other will be inappropriately spaced, with unfortunate consequences for the quality of the image produced. A useful rule of thumb is that the maximum angular resolution is determined by the width of the synthesized beam peak (which in turn is determined mainly by the length of the longest baselines in the array), and the maximum angular size of the radio source that can be imaged is determined by the synthesized beamwidth multiplied by the total range of baseline lengths in the array.

In practice, it is necessary to average the correlator output in both time and frequency, and this also sets a limit on the angular extent of radio emission that an interferometer array can image. Averaging visibilities in time results in an azimuthal smearing of the measurements, resulting in suppression of emission far from the image center. Averaging in frequency has a similar effect in the radial direction (this can be understood as an effective change in baseline length, in wavelengths, as the frequency changes across the bandpass).

Image Deconvolution

The dirty image produced by an interferometer array can be cleaned by deconvolving it with the dirty beam. This removes the effects of the dirty beam's sidelobes and results in an image with higher dynamic range (i.e., weaker features can be detected). There are a number of approaches to the deconvolution process, none of them straightforward because the existence of (u,v) points with zero amplitude prevents deconvolution via a simple division of Fourier transforms.

The two more common deconvolution algorithms in radio astronomy are CLEAN¹⁶ and maximum entropy.^{17,18} The CLEAN approach iteratively subtracts the dirty beam centered on the brightest pixel in the image until no pixels much higher than the noise level remain. During each dirty-beam subtraction, a delta function is placed at the location of the center of the dirty beam. Thus, after the subtractions we have an image consisting of many delta functions located at some of the image pixel centers. This array of delta functions is then convolved with a clean beam to restore an appropriate angular resolution (consistent with the actual (u,v) sampling by the array), but without adding back any sidelobes. This approach is simple, intuitive, and usually produces dramatic improvements in image dynamic range. It can fail to converge in some cases, but usually only on very extended sources. The CLEAN algorithm assumes that the sky can be modeled as a finite number of point sources (the delta functions), which is reasonable for compact radio sources surrounded by empty sky. However, for very extended sources, most of the field of view may contain radio emission, and then the assumption that delta functions are a good approximation can break down.

For extended sources, an algorithm with the opposite assumption of the one for CLEAN is commonly used. This is the maximum entropy algorithm, which assumes that the sky is uniformly bright (featureless) and tries to produce an image as close to featurelessness as the data allow. In other words, this algorithm produces an image with the minimum amount of structure (and thus the maximum entropy) consistent with the visibility data. It tends to image very extended radio emission more successfully than CLEAN does. It also has the interesting feature that the images it produces have higher angular resolution in areas of higher SNR, since no equivalent of a clean beam is used directly. Sometimes it is instructive to use both approaches to image a given set of observations. The differences between images produced using such different starting assumptions can be a useful indicator of the true uncertainties in the image plane.

Multifrequency Synthesis Imaging

It is always desirable to have more complete sampling of the visibility function, but there are limits to the number of telescopes and to the observing time available. A relatively new technique to improve (u,v) sampling is to observe in several separate frequency bands simultaneously. This has the effect of sampling a number of (u,v) points along a radial direction in the (u,v) plane simultaneously for each baseline. If the total fractional frequency range spanned by the observations is comparable to the fractional spacing of baselines in the array, a large reduction in the size of unsampled (u,v) regions can be obtained in this way. Consequently, the synthesized beam will have much lower sidelobes, improving the speed and probably the ultimate quality of image deconvolution. The limitation is that over a wide frequency range, the structure of the radio source could change, which would make it inappropriate to combine all of the visibility data into a single image. In principle, it is possible to solve for the spectral characteristics of the source in addition to other parameters, but this is a significant complication to the imaging process.

Wide-Field Imaging

Throughout this chapter, we have been discussing interferometers and imaging in terms of two-dimensional arrays (the (u,v) plane, the image plane, and two-dimensional Fourier transforms). This is appropriate for the usual applications of radio interferometers, which are normally used to image relatively small regions of the sky and which (except for very-long-baseline arrays) are built with the telescope approximately in a plane.

For imaging large regions of the sky, the two-dimensional assumption is not adequate. Instead, it is necessary to replace the (u,v) plane with a (u,v,w) volume in which the visibility function is measured. Similarly, the image (sky) plane must be replaced by a spherical surface in a three-dimensional image volume. Then, it is also necessary to use a (computationally intensive) three-dimensional transform to move between these volumes.

A number of techniques have been developed to avoid the need for three-dimensional transforms in wide-field imaging, the most common of which involves dividing the field of view into many smaller regions that can each be approximated as two-dimensional. A newer and more efficient approach developed by T. Cornwell¹⁹ is to use Fresnel diffraction theory to calculate the visibility sampling that would be obtained if all antennas were in a single plane orthogonal to each source. This approach is called w projection.

Wide-field imaging can also complicate the calibration of interferometer data. Up to now we have assumed that propagation effects (ionospheric and atmospheric attenuation and phase fluctuations) could be represented by a single time-variable error term per telescope. However, emission from a large area of sky could pass through regions of the ionosphere or atmosphere containing different, uncorrelated conditions. This implies that we need multiple simultaneous direction-dependent error terms per telescope. The number of error terms could now exceed the number of available visibility measurements, violating the basis for self-calibration.²⁰ These are areas of active research, and new approaches are likely to be offered within the next few years.

49.5 MAJOR RADIO TELESCOPES

Single-Aperture Telescopes

This section gives a brief description of a few of the world's great single-aperture telescopes.

Arecibo Radio Telescope The Arecibo Observatory (Puerto Rico) is one of the world's largest radio telescopes.²¹ Its main reflector is a spherical surface 305 m in diameter and nearly 51 m deep. Forty thousand perforated aluminum panels compose the surface of the telescope. The panels are supported by a network of steel cables strung underneath them. The Arecibo feed system consists of both narrowband line feeds and Gregorian optics to correct for spherical aberrations, as shown in Figure 49-4. The feed system is suspended 137 m above the reflector and attached to a triangular frame supported by cables strung from three towers. Immediately below the frame, a 100-m-long azimuth arm moves on a circular track. The curved part of the azimuth arm is another track. On one side of this track is a carriage house that supports

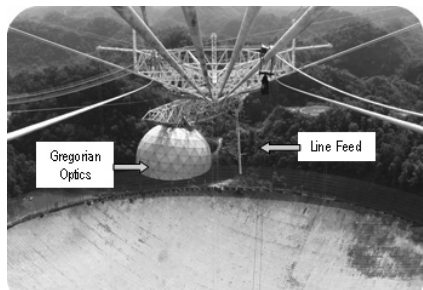


FIGURE 49-4 Arecibo feed system

the narrowband line feeds; on the other side, a large dome houses the optics. Inside the dome, two subreflectors (secondary and tertiary) direct radiation to a set of wide-band horns. Either the Gregorian optics or line feeds can be positioned up to 20 deg from the vertical.

Green Bank Telescope The largest clear-aperture telescope is the 100-m-diameter Robert C. Byrd Green Bank Telescope (GBT) in West Virginia.²² A clear aperture is significant because clear apertures prevent reflection and diffraction from the telescope's feed support structure. The clear-aperture design was achieved by implementing an offset reflector geometry that could be used in a prime-focus mode or a dual-reflector Gregorian mode. The main reflector is composed of 2004 panels mounted at their corners on actuators, which are small motor-driven pistons that enable a surface adjustment for both thermal- and gravity-induced surface distortions. The antenna's wheel-and-track design allows it to scan the entire sky above 5 deg of elevation. The track measures 64 m in diameter; to ensure precise pointing, it is level to less than .01 cm.

Effelsberg Radio Telescope Until the Green Bank Telescope was built, the largest fully steerable telescope was the 100-m-diameter reflector in Effelsberg (Germany), which began operating in 1972.²³ The telescope uses dual-reflector Gregorian optics at both the primary and secondary focus to observe radio emission in the wavelengths from 90 cm down to 3.5 mm. Because of the homology of the design and a special support structure, the deviation of the dish surface from the ideal parabolic form is at most 0.5 mm despite the flexing of the steel members of up to 10 cm due to gravity. The shift in the focus position due to surface flexure accompanying the tilting of the dish is compensated for by means of an electronic control mechanism.

The receiver systems are mounted either at the focal point of the main reflector, just beneath the prime-focus cabin suspended on four support legs, or at the secondary focus located at a central point just above the surface. There, in the secondary focus cabin, it is possible to use many additional receiver systems, some even simultaneously.

Parkes Radio Telescope The 64-m Parkes Radio Telescope (Australia) began operations in 1961.²⁴ It was then one of the first large movable dishes in the world. Today, Parkes frequently operates together with the Australia Telescope Compact Array at Narrabri and a single dish at Mopra, to form a very-long-baseline interferometry array.

The original surface of the telescope was metal mesh but was later upgraded to smooth metal panels on the central part of the antenna, allowing for operation at both centimeter and millimeter wavelengths (see Figure 49-5). The receiving cabin is located at the focus of the parabolic dish, supported by three struts 27 m above the dish. The cabin contains multiple feed systems, which can be switched into the focal point for different science observations.

NASA applied the basic design concept of the Parkes Radio Telescope when developing the Deep Space Network. Subsequently, almost identical 64-m antennas were built at Goldstone (California), Madrid (Spain), and Tidbinbilla (Australia). In 1987, NASA extended the Tidbinbilla antenna from 64 m to 70 m, surpassing Parkes as the largest steerable radio telescope in the Southern Hemisphere.

Major Interferometers

This section gives a brief description of three of the world's largest interferometers.

The Very Large Array The Very Large Array (VLA) consists of 27 reflector antennas in a Y-shaped configuration located 80 km west of Socorro, New Mexico.²⁵ Each antenna is



FIGURE 49-5 The Parkes Radio Telescope

25 m in diameter. Each reflector has multiple narrowband feed systems covering the range of frequencies from 70 MHz to 43 GHz. The elements can be respaced in their Y pattern in four configurations: A array, with a maximum antenna separation of 36 km (which has the sensitivity of a 130-m antenna); B array, with a maximum separation of 10 km; C array of 3.6 km; and D array of 1 km. The telescopes are switched between these configurations approximately every four months. The resolution of the VLA is set by the size of the array—up to 36 km across. At the highest frequency, resolution is 0.04 arcsec.

With a view to upgrading the 1970s technology used to build the VLA, recent funding has been allocated for the conversion of the VLA into the Expanded Very Large Array (EVLA). The upgrade will enhance the instrument's sensitivity and frequency range by installing new hardware at the San Agustin site.

Giant Metrowave Radio Telescope The Giant Metrowave Radio Telescope (GMRT) was established by the National Centre for Radio Astrophysics (NCRA) of India to do research using the meter wavelengths range.²⁶ The site is about 80 km north of Pune, India. GMRT consists of 30 fully steerable 45-m-diameter parabolic antennas. The main reflector is lightweight and of low solidity because the conventional backup structure was replaced with a series of rope trusses (made of thin stainless steel wire) stretched between 16 parabolic frames of tubular steel. The wire ropes are tensioned suitably to create and maintain a mosaic of plane facets approximating a parabolic surface. A lightweight thin wire mesh (made of 0.55-mm-diameter stainless steel wire) with a grid size varying from 10×10 mm in the central part of the dish to 20×20 mm in the outer parts, and stretched over the rope truss facets, forms the reflecting surface of the dish. Making the reflecting surface of a porous mesh cuts down the wind forces by a large factor and is particularly suited to the local climate conditions of the plains.

The number and configuration of the dishes were optimized to achieve sensitivity at high angular resolution as well as the ability to image radio emission from diffuse extended regions. Fourteen of the 30 dishes are located randomly in a compact central array in a region about 1 km². The remaining 16 antennas are spread out along the three arms of an approximately Y-shaped configuration over a much larger region, with the longest interferometric baseline at about 25 km.

The array operates in six frequency bands centered at 50, 153, 233, 325, 610, and 1420 MHz. All feeds provide for dual polarization outputs and, in some configurations, dual-frequency observations are possible.

The Atacama Large Millimeter Array The Atacama Large Millimeter Array (ALMA), in Chile, is a millimeter-wavelength telescope planned to consist of about sixty-four 12-m antennas located at an elevation of 5000 m.²⁷ ALMA functions at between 10 mm and 350 μ m, with array configurations from approximately 150 m to 10 km. The spatial resolution is 10 milliarcsec; ALMA will be able to image sources arcminutes to degrees across at 1-arcsec resolution.

An important requirement for the ALMA project was the capability to rapidly switch positions in order to achieve phase-stable observations a large fraction of the time. To achieve this goal, in 1992, work began on a “slant-axis” antenna design, whose stiffness enabled very fast switching.

Later, the slant-axis design evolved into an off-axis paraboloid whose advantages included good gravitational performance, lack of counterweights, and the high sensitivity of an unblocked aperture. However, a minimum-blockage conventional antenna design, patterned after the Berkeley Illinois Maryland Association (BIMA) 6-m antennas (in California), proved to be more cost effective and provided better field-of-view and polarization performance and was the design chosen for development.

In 1998, ALMA was formed as an international partnership consisting of both the North America and European radio astronomy communities. Each partner had its own antenna design concepts. The international partnership decided that the best way to determine the antenna design would be for each organization to produce a prototype antenna for testing (see Figure 49-6). In 2001, the Japanese joined the ALMA partnership and offered a third antenna design for testing. The three prototypes were tested at the VLA site; all three designs were selected for use in the ALMA array.

49.6 CONCLUSIONS

Radio telescope antennas have produced dazzling science covering meter-to-submillimeter wavelengths. The next generation of telescopes such as ALMA, the Square Kilometer Array (SKA),²⁸ Low-Frequency Array (LOFAR),²⁹ and the Allen Telescope Array (ATA)³⁰ promise even more wondrous science. Perhaps, with the ATA or SKA, we may even discover that we are not alone in the universe.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.



FIGURE 49-6 Prototype ALMA antennas being tested at Socorro, New Mexico

REFERENCES

1. M. Ryle and D. Vonberg, "Solar Radiation on 175Mc/s," *Nature*, vol. 158 (1946): 339.
2. J. Pawsey, R. Payne-Scott, and L. McCready, "Solar Radiation at Radio Frequencies and Its Relation to Sunspots," *Nature*, vol. 157 (1946): 158.
3. A. S. Bennett, "The Revised 3C Catalogue of Radio Sources," *Mem. Royal Astron. Soc.*, vol. 68, no. 163 (1962).
4. S. Chandrasekhar, *Radiative Transfer* (Mineola, New York: Dover Publications, 1960): 24.
5. A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy* (New York: Wiley-Interscience, 1986).
6. E. M. Leitch, J. M. Kovac, N. W. Halverson, J. E. Carlstrom, C. Pryke, and W. E. Smith, "Degree Angular Scale Interferometer 3 Year Cosmic Microwave Background Polarization Results," *Astrophys. J.*, vol. 624, no. 10 (2005): 10–20.
7. F. Piacentini, P. Ade, J. Bock, J. Bond, J. Borrill, A. Boscaleri, P. Cabella, et al., "A Measurement of the Polarization-Temperature Angular Cross-Power Spectrum of the Cosmic Microwave Background from the 2003 Flight of BOOMERANG," *Astrophys. J.*, vol. 647 (2006): 833.
8. R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd Ed. (New York: McGraw-Hill, 1978).
9. H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *Trans. Am. Inst. Electr. Eng.*, vol. 47 (1928): 617.
10. P. J. Napier, A. R. Thompson, and R. D. Ekers, "The Very Large Array: Design and Performance of a Modern Synthesis Radio Telescope," *Proc. IEEE*, vol. 71 (1983): 1295–1320.
11. A. R. Thompson, B. G. Clark, C. M. Wade, and P. J. Napier, "The Very Large Array," *Astrophys. J. Suppl.*, vol. 44 (1980): 151–167.
12. J. H. Van Vleck and D. Middleton, "The Spectrum of Clipped Noise," *Proc. IEEE*, vol. 54, no. 2, (January 1966): 2–19.

13. NASA, *Radio Interferometry Techniques for Geodesy*, NASA Conference Publication 2115, (Washington, DC: National Aeronautics and Space Administration, 1980).
14. T. J. Cornwell, "Self Calibration," *NRAO Workshop Proceedings*, vol. 5, A. R. Thompson and L. R. D'Addario, (eds.) (Green Bank, WV: National Radio Astronomy Observatory, 1982).
15. T. J. Pearson and A. C. S. Readhead, "Image Formation by Self Calibration in Radio Astronomy," *Annual Rev. of Astron. and Astrophys.*, vol. 22 (1984): 97–130.
16. J. A. Hogbom, "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines," *Astron. Astrophys. Suppl.*, vol. 15 (1974): 417.
17. J. G. Ables, "Maximum Entropy Spectral Analysis," *Astron. Astrophys. Suppl.*, vol. 15 (1974): 383.
18. S. F. Gull and G. J. Daniell, "The Maximum Entropy Method," in *Image Formation from Coherence Functions in Astronomy*, C. van Schooneveld, (ed.) (Amsterdam: Reidel, Dordrecht, 1979): 219.
19. T. J. Cornwell, K. Golap, and S. Bhatnagar, "W Projection: A New Algorithm for Non-Coplanar Baselines," EVLA Memo 67, NRAO, 2003.
20. F. R. Schwab, "Relaxing the Isoplanatism Assumption in Self-calibration: Applications to Low-frequency Radio Interferometry," *Astron. J.*, vol. 89 (1984): 1076.
21. National Astronomy and Ionosphere Center, <http://www.naic.edu/> (accessed Oct. 31, 2006).
22. National Radio Astronomy Observatory, <http://www.gb.nrao.edu/> (accessed Oct. 31, 2006).
23. The Effelsberg 100-m Radio Telescope, <http://www.mpifr-bonn.mpg.de/div/effelsberg/> (accessed Oct. 31, 2006).
24. Parkes Observatory, <http://www.parkes.atnf.csiro.au/> (accessed Oct. 31, 2006).
25. National Radio Astronomy Observatory Very Large Array, <http://www.vla.nrao.edu/> (accessed Oct. 31, 2006).
26. Giant Metrewave Radio Telescope (GMRT), <http://www.gmrt.ncra.tifr.res.in/> (accessed Oct. 31, 2006).
27. Atacama Large Millimeter/Submillimeter Array, <http://www.alma.nrao.edu> (accessed Oct. 31, 2006).
28. SKA (Square Kilometre Array), <http://www.skatelescope.org/> (accessed Oct. 31, 2006).
29. LOFAR (Low Frequency Array), <http://www.lofar.org/> (accessed Oct. 31, 2006).
30. Allen Telescope Array, <http://ral.berkeley.edu/ata/> (accessed Oct. 31, 2006).

BIBLIOGRAPHY

- Christiansen, W. N. and J. A. Hogbom, *Radio Telescopes* (Cambridge: Cambridge University Press, 1985).
- Meeks, M. L., (ed.) *Methods of Experimental Physics*, vol. 12, parts B and C (New York: Academic Press, 1976).
- Thompson, A. R., J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy*, (New York: Wiley-Interscience, 1986).

Chapter 50

Indoor Antenna Measurements

Walter D. Burnside
Inder J. Gupta
Teh-Hong Lee

The Ohio State University

CONTENTS

50.1 INTRODUCTION	50-2
50.2 IMPEDANCE OR S11 MEASUREMENTS	50-3
50.3 PATTERN OR S12 MEASUREMENTS	50-4
50.4 ABSORBER	50-11
50.5 TAPERED R-CARD FENCES	50-14
50.6 HARDWARE-IN-THE-LOOP CHAMBERS	50-15
50.7 GAIN, EFFICIENCY, AND BORESIGHT MEASUREMENTS	50-17
50.8 POLARIZATION MEASUREMENTS	50-19
50.9 ANTENNA TEMPERATURE MEASUREMENTS	50-20
50.10 CHAMBER EVALUATION	50-21
50.11 CLUTTERED ENVIRONMENT MEASUREMENTS	50-24
50.12 SUMMARY AND CONCLUSIONS	50-25

50.1 INTRODUCTION

Over the last few decades, there has been great progress in the development of indoor measurement facilities; and as a result, many outdoor facilities are closing in lieu of going indoors. This has been made possible because of the development of enhanced measurement hardware and facilities. This chapter attempts to describe how one can take advantage of these modern measurement systems in order to perform various types of antenna measurements.

There have been major enhancements in terms of network analyzers such that one is now able to quickly and easily collect both impedance and pattern data using these very powerful devices. These measurements are normally done in CW mode, but more and more applications have increased bandwidth such that time domain measurements can also be taken with most modern network analyzers. This can be done using the time domain option in the network analyzer by acquiring broadband (amplitude and phase) data and then transforming from the frequency to the time domain. Once in the time domain, it's possible to software time gate the desired time response and, therefore, gate out the undesired time terms. This will be shown to be very useful later, when the various antenna measurements are considered in more detail.

Besides using a network analyzer, one can also consider many other source and receiver systems. These options include time domain hardware that can be used to directly gate out undesired time responses using pulse hardware. There are two rather inexpensive hardware options that are considered here: 1) a *chirp system* and 2) a *pulsed/CW system*.

The *chirp system* uses a linear frequency sweep for the transmitted signal. This chirp signal is received by the antenna-under-test (AUT), where it is then sent to a mixer that correlates it with a sample of the linear chirp coming directly from the source. The output of this mixer is then a low-frequency harmonic that is directly related to the difference in frequency between the AUT received signal and the present frequency of the swept source. To isolate the desired frequency harmonic, simply filter the desired AUT term. These systems are rather inexpensive because they consist of a swept source, a mixer, and a low-frequency digital filter system for selecting the desired frequency response. In order to properly use a chirp system, one must understand and maintain the amplitude and phase linearity requirements and must also create the proper amplitude weighting to control range sidelobes.

An inexpensive *pulsed/CW system* can be built using a network analyzer and pulse hardware. Just about any network analyzer can be considered for use with this application. The pulse hardware consists of a single-pull-single-throw (SPST) high-speed transmit switch, which creates the pulse waveform, and a power amplifier to increase the radiated signal. This hardware goes between the network analyzer transmit port and source antenna. At the AUT, which is considered as the receiver in this example, one then adds another SPST high-speed switch to select the desired time response. This received pulse signal is then input to the network analyzer using the receive port. Because this is a broadband signal that consists of pulsed harmonics, one has to be careful to choose a pulse repetition rate so that pulse harmonics are not related to the receiver harmonics. Otherwise, the network analyzer received response will be corrupted by the pulse harmonics. Recall that the network analyzer is used to transmit and receive a CW signal. So, the pulsed/CW hardware is used to time gate out undesired time responses, while maintaining the CW response as seen by the network analyzer. In designing these systems, one has to be very careful to avoid ringing within the pulsed hardware.

In times past, one had to worry about the transceiver hardware as much as the facility issues associated with his or her measurement. As stated earlier, that is no longer the case. Modern network analyzer and transceiver systems are extremely linear, have excellent IQ circularity, are very stable, and provide huge dynamic ranges. As a result, modern antenna measurements can now focus on the facility capability as opposed to the hardware that is used to collect the desired data. These same conclusions were pointed out by Dybdal.¹

Since indoor facilities are limited to reasonable dimensions, the effective distance between the transmitter and receiver is not that large. That being the case, there is not a need to transmit much power. Thus, the normal output provided by a network analyzer is usually sufficient for most indoor antenna measurements. If more power is needed, one does not have to typically use more than a solid-state amplifier, which is very nice because solid-state units are readily-available, reasonably priced, and stable.

Before going into the details of the different facilities, let's think about the topic in general. First, any antenna measurement should be done in such a way that the actual antenna application is simulated as well as can be accomplished within the confines of an indoor environment. For example, one cannot simulate a full satellite antenna measurement including all the associated satellite features. That can only be done within certain limits. Second, the indoor environment creates undesired scattering that must be understood and controlled to the degree necessary for meeting the desired measurement goals. These two issues tend to mean that the facility will be extremely large. Obviously, on the other side, building costs force the size the facility to be smaller. As a result, one has to make compromises between cost and performance. That being the case, present-day facilities are not normally perfectly designed for any specific measurement; therefore, they must be evaluated to determine how well they can actually perform any given measurement.

Since modern antenna measurement facilities are not ideal for any given measurement, they must be carefully thought out before they are even purchased in the first place. Furthermore, the burden of properly selecting and designing these facilities has to be carried out in such a way that the facility procurer plays a very active role along with the vendors. Therefore, it's important to get smarter in terms of measurement facilities. In that regard, one should not think of the measurement as a separate and distinct task when compared to numerically simulating the antenna. If one understands the full volumetric pattern of the antenna, it's possible to predict a measured response for any known complex signal illuminating the AUT. This being the case, one can determine the specific fields in the test zone by collecting high-quality probe data. These probe fields can then be transformed to determine the incident field plane wave spectrum, which describes the actual illumination of the AUT within the facility. With this information, one can then take this incident plane wave spectrum and compute the expected measured results that are achievable within a specific facility. If these simulated measured results show that the resulting error exceeds the desired measurement goals, then that facility is not appropriate for the proposed AUT tests. In fact, this concept should be adopted as a standard practice for any proposed antenna measurement in order to justify that a given facility can actually meet or exceed the desired measurement goals. More will be said about this topic later.

To become more familiar with these measurements or to get a better basic understanding of them, refer to the IEEE Standard Test Procedures for Antennas.² This chapter can then be used to bring these concepts up-to-date. Note that each of the measurement topics are briefly described in the following sections along with appropriate references to allow the reader to examine this material in as much detail as desired.

50.2 IMPEDANCE OR S11 MEASUREMENTS

As stated in the introduction, impedance or S11 measurements have been greatly improved with the advent of modern network analyzers and the development of accurate loads needed for proper calibration. Using the calibration procedures fully described in the network analyzer documentation, one can easily achieve precise impedance results provided that an

appropriate antenna environment for the measurement is used. In other words, the measurement hardware is typically not the limiting factor in these measurements. It is the operator and the measurement environment that can limit S11 measurement performance.

In order to understand how well an S11 measurement is performed, one should use the time domain option. This is very useful because the time domain can be used to ascertain the amplitude and placement of the various time responses relative to the desired AUT response. If the network analyzer calibration is correctly performed, the time responses before the AUT response should be very small compared to the AUT S11 term since this response involves cabling and connectors. The environment can be evaluated in terms of the time responses after the AUT term or terms. If all the time responses, before and after the AUT term or terms, are negligibly small compared to the AUT responses, then the calibration should be correct and the AUT environment is sufficiently controlled for a successful S11 measurement. Therefore by deleting these early and late time responses through time gating, the gated response that has been transformed into the frequency domain represents the desired AUT response, except near the frequency end bands, which are corrupted by the time gate. Therefore, one should use a wider bandwidth than is needed for the actual measurements so that the end band corrupted values do not impact the desired data.

Recall that a gain measurement requires a very accurate calibration relative to some type of gain standard. Since one can perform very accurate impedance calibrations, it is very interesting to think of ways to take gain measurements using an S11 measurement. Such a concept will be presented in the section discussing gain.

50.3 *PATTERN OR S12 MEASUREMENTS*

For pattern measurements, one can again use the various measurement hardware systems described in the introduction. These systems will all perform accurate pattern measurements, provided that one can achieve the desired signal-to-noise ratio needed for the measurements. Recall that antenna patterns can have a huge dynamic range, especially for low-sidelobe antenna systems. Even so, the hardware can typically achieve a dynamic range of say 80 dB as long as the peak pattern level is adjusted so that it is close to the saturation level of the receiver system. Across this 80 dB dynamic range, the measurement should be very linear and provide excellent IQ performance. If more dynamic range is needed, the pattern should then be taken in terms of two patterns with the second pattern taken with more amplification added in order to measure the lower levels. Note that low-noise solid-state amplifiers should typically be used in this situation. In these cases, a preliminary link budget and attention to the receiver noise floor and saturation is very useful.

Clearly with antenna patterns, one must deal with an antenna motion control system. This issue is oftentimes implemented in terms of massive hardware that is needed to meet or exceed some rotational accuracy requirements. These requirements tend to get very convoluted in the sense that one defines the maximum weight and angular accuracy together. This places a very stiff set of specifications on the whole rotational system independent of the measurement being taken. For example, there are facilities that take low-gain horn pattern measurements using massive elevation-over-azimuth rotational hardware. In these cases, it is often difficult to see the small antenna in comparison to the very large rotators and associated support structure. So what can be done to correct this undesirable situation? This will be examined next.

One must remember that the antenna pattern should be measured as though it is mounted in its natural environment, which in most cases means that it is surrounded by free space. Thus, the rotation hardware has to be configured based on the antenna being measured. One must use a small low-scattering mount for low-gain light-weight antennas, whereas

large metal mounts should be used for high-gain heavy antenna systems. That being the case, the rotational hardware has to be designed to be multifunctional. In fact, a high-quality antenna measurement facility needs to have different rotational hardware that has been specifically developed to treat small, medium and large antenna systems. It is interesting to note here that the rotational hardware used for a small antenna obviously does not need the rotational accuracy that a high-gain antenna needs. This is very important because it implies that one can use a much smaller mount and encoder and yet achieve or exceed the necessary rotational motion control and accuracy.

For cases where one wishes to measure the complete volumetric pattern for a low-gain antenna, it is important to use a mount structure that is virtually transparent. This can be done using a foam column sitting on an azimuth rotator, keeping in mind that the cable remains as a potential scatterer. To minimize the impact of the cable, one should consider a fiber-optic cable solution. In any event, one can collect a full azimuth cut with very little impact caused by the foam column as compared to a massive elevation-over-azimuth approach. However, one cannot easily collect a full volumetric pattern. To overcome this limitation, one can add a small roll axis rotator on top of the foam column. This will result in a roll-over-azimuth system that can be used to collect the full volumetric pattern for low-gain light-weight antennas. For medium sized antennas, one can replace the foam column with a light-weight foam and plastic pole. Then, a medium-sized roll axis rotator can be mounted on top of this pole to again provide roll-over-azimuth motion. Thus, it is specifically-designed for medium gain and weight antennas. For the largest antennas, one can use the massive rotational equipment stated earlier. This way, one is able to take measurements using a very versatile set of rotational hardware that can be effectively used for all types of antenna measurements.

In order to achieve a given rotational step size, one must realize that the encoder must provide much more resolution than specified by the angular pattern step size. For example, let's suppose that one wishes to measure a low-gain antenna pattern every 1 degree. The encoder cannot have a resolution of 1 degree. Its resolution should be three to ten times finer or 0.3 to 0.1 degrees. This way, the pattern control software can start to watch the encoder readout until the encoder reaches the desired angular position to within its resolution. So the better the resolution, the better and more uniform the actual angular steps will be. Once at the desired angle, the receiver system will then collect the measured data, which will be stored in the computer. The software will then repeat this process until the whole pattern is measured. Putting it another way, if the control software does not have an angular readout resolution finer than the desired pattern accuracy, the precision of the data collection will be very poor. This can cause a problem if one needs precise IQ measured data versus angle. In fact, it will be shown that precise angle data is necessary if one wishes to do image processing of the measured data and is collecting data at the sampling theorem rate. If interested in learning more about antenna pattern image processing, refer to Lee and Burnside.³

So, it has been shown here that one can properly design the rotational hardware to obtain the various antenna pattern measurements performed indoors today. That being the case, one has to now focus on the pattern measurement facility. Since there are a wide variety of such facilities, each one will be examined separately.

Traditional Spherical Range

Traditional spherical ranges consist of a source antenna usually mounted in the transmit wall of a rectangular chamber. The source directly illuminates the AUT mounted in the center of the test zone. The size of this chamber is determined based on the size of the test zone and the frequency of operation. The distance between the source and the center of the test zone is normally defined as $R = 2D^2/\lambda$, where D is the maximum diameter

of the AUT and λ is the wavelength. Note that this requirement means that there will be a 22.5 degree phase taper across the test zone for the frequency that was used to define the wavelength in the definition of R . Clearly, this definition has come from an attempt to approximate a plane wave using a spherical radiator as the illuminator of the AUT. The range (R) must also be several wavelengths long to avoid inductive coupling between the source and AUT. This is the limiting factor at low frequencies for these designs.

The spherical range chamber dimensions are shown in Figure 50-1. The length of the chamber is $L = R + X + D/2$, where D is again the diameter of the test zone. The width and height of the chamber is $W = 2X + D$. The dimension is $X = D$ provided that D is greater than the maximum wavelength across the full operational bandwidth of the chamber. If the maximum wavelength is larger than D , then X is defined as the maximum wavelength.

Note that all these dimensions are measured inside the absorber and represent an initial chamber layout guideline. The absorber will be studied in a later section.

This type of chamber is normally used to measure the antenna patterns for low to medium gain antennas. The test zone size is obviously a function of frequency. The diameter of the test zone versus frequency is given by $D = \sqrt{R * \lambda / 2}$, provided that one wishes to simulate plane wave illumination. Thus, the test zone size reduces with increasing frequency, which means that this type of chamber is not really appropriate for high gain antenna pattern measurements.

One of the major issues associated with the spherical range is that one tends to build them as long narrow tunnels in order make the range (R) as long as possible. This is very problematic as will be seen in the section discussing absorbers. As the chamber becomes long and narrow, the source will tend to illuminate the absorber near grazing incidence. As one would expect, the absorber reflection level is very sensitive to the angle-of-incidence and in fact approaches unity at grazing incidence. Consequently, one should always question using an absorber when the angle-of-incidence is approaching grazing incidence. It is best to think of the absorber as working well for normal incidence or 0 degrees to no more than about 50 degrees. With this 50 degree angle-of-incidence limitation for absorber, one cannot think of making a spherical range chamber that is really that narrow and long.

To get around this absorber limitation, it's possible to design a spherical chamber using a tapered R-Card fence concept. This type of fence will block the undesired absorber scattering near grazing incidence without creating an undesired fence diffraction term. Therefore, one can make a long narrow spherical range chamber, provided that the appropriate tapered R-Card fence or fences are used. These fences are described in the section discussing tapered R-Card fences.

Tapered Chamber

A tapered chamber is actually a spherical range that is housed in a chamber that looks much like a horn, as shown in Figure 50-2. This chamber works very well provided that the source antenna is located at the throat of the horn section. If that is the case, the source energy simply propagates outward from the source to illuminate the test zone. The dimensions of this chamber follow the same formulas listed earlier for the spherical range chamber.

Until recently, the major limitation associated with the performance of a tapered chamber has been the need for an ideal source antenna. The problem is that the

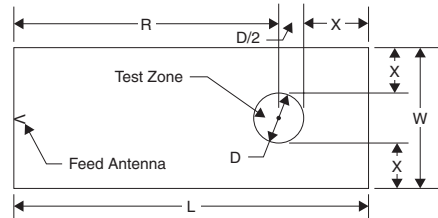


FIGURE 50-1 Spherical range chamber dimensions

source antenna phase center must be located at the vertex of the throat section, which means that the antenna's phase center is surrounded by an absorber. Thus, the absorber tends to attenuate the radiated energy. There have been two antennas that have been specifically-designed to solve this problem: a tapered R-Card horn antenna⁴ and a quad-ridge antenna.⁵

The quad ridge antenna is most appropriate because it is dual-polarized and it can be permanently integrated into the throat section. It also has very good bandwidth properties so that it can typically provide complete performance across the full operational bandwidth of the tapered chamber.

The tapered chamber is used for low and medium gain antenna patterns because of the range (R) requirement. Furthermore, it is most appropriate for low frequency measurements because it has been designed to handle the absorber grazing incidence problem. Recall that the absorber must get thicker for lower frequencies. Thus, a spherical range chamber will grow in size as the frequency is decreased. That is not the case for the tapered chamber in that the absorber is illuminated at grazing incidence; so therefore, it does not have to be very thick. Consequently, one should always think of a tapered chamber as the best approach for low frequency, low and medium gain antenna pattern measurements. One should refer to Burnside et al⁶ to determine the proper layout and absorber treatment for a modern tapered chamber.

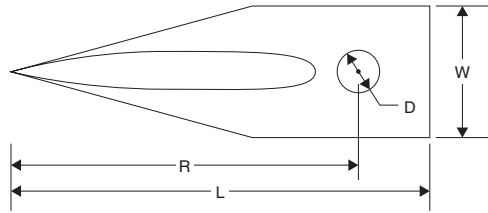


FIGURE 50-2 Tapered chamber dimensions

Compact Range

A *compact range* is used to create a plane wave field in the test zone as presented in "Compact Range Techniques and Measurements."⁷ This is done by using a parabolic reflector with the feed antenna located at its focus. As a result, the reflected field from the parabola that illuminates the test zone typically has a very uniform phase; while, the amplitude varies slightly caused by the feed pattern and range variation related to the distance between the feed antenna and the reflector surface. In this case, the test zone lies within the near field of the compact range reflector providing two advantages. The first advantage is that the test zone fields are not spherically attenuated so that the power density in the test field for a given source power is increased enhancing measurement dynamic range. The second advantage is that the collimated near field of the compact range reflector minimizes the illumination of the facility's side walls, floor, and ceiling. As a result, one finds for well-designed compact ranges that the amplitude taper is about 1 dB, the amplitude ripple is about one tenth of 0.1 dB, and the phase variation is typically less than 6 degrees even at its highest operational frequency.⁸

The major issue associated with a well-designed compact range is again the issue of stray signals. As stated earlier, the reflected field creates the desired illumination of the test zone. However, the reflecting surface must be terminated properly to fit within a finite chamber. There have been many attempts to find the best compact range reflector termination concept. This challenge has been going on for more than 20 years and everyone has learned a lot about the different approaches since all of them have been built in one form or another. The blended rolled edge reflector, as designed in Gupta et al,⁹ clearly performs better than any other approach but it is also the most expensive. For example, it is not unusual for a blended rolled edge compact range to have no more than a few tenths of a dB ripple in the test zone fields. This implies that the stray signal levels are nearly 40 dB below the desired plane wave illumination.

Let's suppose that one wishes to measure a low sidelobe antenna. It means that when the main beam of the AUT points toward the stray signal direction, it will be boresighted with the stray signal term. Consequently, the main beam gain will amplify a weak stray signal and create an undesired sidelobe in the resulting measured pattern. This is an issue for all reflector termination designs, but the rolled-edge reflector will still provide about 15 dB better performance. The full impact is actually more significant than just the 15 dB. For serrated edge reflectors, the serrations typically cover a significant portion of the overall reflector surface. So, their stray signals are seen from the AUT point-of-view for a relatively wide angular region. The blended-rolled edge reflector discontinuity comes from the junction between the parabola and blended-rolled edge. This junction is normally designed to be about halfway from the center of the reflector to outside rolled edge; therefore, it has much less angular impact on a pattern measurement. If one is interested in a very thorough discussion of the differences between blended-rolled edge and serrated-edge reflectors, refer to Lee and Burnside.¹⁰⁻¹¹

All compact range systems have another significant stray signal that must be properly handled. This is the feed spillover into the test zone. Since this undesired term takes a much shorter path than the desired plane wave, it can be easily time gated out. Thus, it is important to use some type of time domain technique to minimize the impact of this undesired error term. There have been many attempts to remove this feed spillover term by using various types of absorber panels that hide the feed antenna as seen from the test zone. This is not a good approach because it adversely blocks the wavefront coming from the reflector. This, in turn, results in additional scattering that cannot be gated out because it has nearly the same time delay as the desired plane wave term.

Beside these issues, it has been noted that absorber fences have been used in many compact range applications to hide the large antenna mount structures and to provide a convenient access area to insert or remove the AUT. These fences are typically a simple plywood structure to which absorber panels are added and tend to run across the full width of the chamber. This seems most appropriate for production applications because one can quickly mount or dismount the AUT using massive rotational hardware. Unfortunately, these fences cause very serious stray signal errors that come from the top edge of the fence. Since the absorber does its job, it creates a very significant shadow directly behind the fence. This shadow region is very similar to the one caused by a metal plate. As a result, the fence edge diffraction can be thought of as coming from the incident shadow boundary associated with a metal plate. With this in mind, the fence causes a stray signal that can severely degrade the test zone fields and ultimately the desired antenna pattern measurements. So, it is better not to use a fence at all. But if a fence is absolutely necessary, then one should use a tapered R-Card fence as described in the tapered R-Card fence section.

A compact range reflector system can be defined in a generic sense as shown in Figure 50-3. The reflector focal distance is defined as F , the reflector size is H , the range from the reflector to the center of the test zone is R , and the test zone diameter is D . Typically, one finds that $H = 2D$, $F = H/2 = D$, and $R = 2F = 2D$. Thus, the compact range reflector system can initially be approximated based on the size of the test zone.

Now that the reflector system has been roughly laid out, one can start to define the chamber dimensions that are needed to house this compact range reflector system. The chamber geometry is shown in Figure 50-4. The length of the chamber is approximately given by $L = D/2 + R + 3D/2 = R + 2D = 6D$.

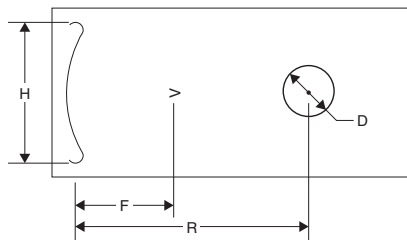


FIGURE 50-3 Compact range system dimensions

The width and height of the chamber are roughly given by $W = 3D$. Again, these are only meant as initial layout dimensions to get the whole design process started. Even so, they are very valuable in terms of estimating initial costs. Furthermore, these dimensions do not include the absorber.

The compact range is the best approach to create a plane wave field across a large test zone, because the reflector is basically acting as a lens to transform an incident spherical wavefront into a reflected plane wave. This planar wavefront is about half the size of the reflector, so it is many wavelengths across. Consequently, it is ideal for measuring the patterns of high gain antennas. Remember that the reflector should be at least 10 wavelengths across in order to actually act as a proper lens.

Because the compact range is complementary to a tapered chamber, many organizations have a tapered chamber to measure low frequency, low to medium gain antennas and a compact range to measure higher frequency, high gain antenna patterns. It is very typical to use a tapered chamber below 1 GHz because of far field limitations and a compact range above 1 GHz. Of course, one can design a compact range to function well for much lower frequencies than 1 GHz and a tapered chamber that works reasonably well above 1 GHz. Nevertheless, the 1 GHz crossover is very typically found in practice.

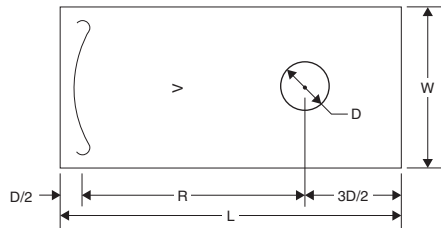


FIGURE 50-4 Compact range chamber dimensions

Near Field Scanner Systems

A near field scanner measurement system is used to collect near field probe data in the close vicinity of the AUT. This probe data is then used to predict a user-desired field result such as far field patterns through the transform between the fields on a surface and the far field values. Each of these systems collects sufficient data that satisfies the sampling theorem for that particular measurement scenario. There are three major near field scanner systems that are commonly used today: 1) planar, 2) cylindrical, and 3) spherical. An overview of these near field systems is presented in Yaghjian.¹²

The planar scanner collects its data across a planar scan region; thus, it is used mainly for radiators that are designed to create a plane wave in the near field, which is indicative of a high gain antenna. The cylindrical scanner is used to collect the near fields over a cylindrical surface; thus, it is most appropriate for cylindrical radiators, such as cell phone basestation antennas that are long and narrow. The spherical scanner is used to collect near field data over a spherical surface; thus, it is most appropriate for low gain radiators that tend to create a spherical phase front close to the AUT. Therefore, the scanner system used for a given measurement is selected based on the radiation characteristics of the AUT. Of course, these systems can be used for a wide variety of applications even beyond the preferred applications for which it is most appropriate. Because of the versatility of these systems, many papers have been written that compare measured results using these different scanners (see Hess,¹³ for example).

These systems have been very carefully evaluated to ascertain the quality of the collected data. They have typically been characterized in terms of an uncertainty table and a resultant measurement uncertainty such as shown in Newell.¹⁴ This table contains all the expected error mechanisms that can occur during a scanner measurement. For each of these error mechanisms, a special test is defined that allows the system designer to determine the impact of that specific error mechanism on a given measurement. Once this is done, each of the error terms is assumed to be random such that they can be summed as random errors. In other words, they are not

coherently related to one another. In fact, it is assumed in this table that each term will have a random error impact during a given measurement much like the impact of noise.

As might be expected, most error terms in the uncertainty table are very small. However, two error mechanisms tend to dominate these uncertainty tables. These two terms are the chamber scattering and probe mutual coupling. The chamber error comes from the fact that the scanner is surrounded by some type of finite chamber. As a result, the measurement will contain chamber effects such as the scattering from the absorber walls. The probe mutual coupling is related to the interaction between the probe used to collect the near field data and the AUT. This term is typically evaluated by moving the probe antenna a quarter wavelength farther away from the AUT and then repeating the measurement. As a result, the AUT mutual coupling term coming back to the probe antenna will change phase by about 180 degrees for the second measurement. This phase change will cause the received signal to be composed of the desired response minus the mutual coupling error term. Thus, the difference between these two measurements tends to isolate this mutual coupling mechanism. Of course, this argument is not exactly true, but it does provide some useful results that give one at least an order of magnitude improvement in this error term. Using a similar approach, the chamber errors can also be studied by moving the measurement setup relative to a given scattering term that one wishes to investigate. At high frequencies, leakage components (waveguide flanges, mixers and so on) can also be troublesome and attention to possible leakage sources is required. In any event, this type of information is used to form an uncertainty error bound for the scanner system. That being the case, one can then use this error bound to determine if a given scanner system can meet the objectives of a given measurement or not.

The planar scanner system is the most popular and has been greatly enhanced over the years.¹⁵ Because the uncertainty table defines the measurement error terms and its associated impact on a given measurement, it has been used to investigate various ways to enhance its performance. This has resulted in very powerful systems that can scan in extremely precise planes at very precise increments even at high speeds and for wide bandwidth applications. However since the scan plane is finite, the resulting pattern is normally used to simply determine the main beam and first few sidelobes. In many cases, the main beam and first few sidelobes are all that is needed; thus, the planar scanner is ideal for these applications. On the down side, all of the scan data must be collected across the full scan plane before the desired far field pattern results can be computed.

The planar scanner is a very important solution for antenna systems that cannot be moved, such as very expensive satellite structures that are composed of multiple antenna systems. In this case, the planar scanner can be placed in front of satellite radiators and then used to probe the near field of each radiating system without moving the satellite structure. With these results, it's possible to determine the far field patterns of each radiating system and verify the boresighting of each system relative to the others. This boresight data can then be related back to the satellite coordinate system to insure that all the radiating systems are radiating properly.

The cylindrical scanner is mainly used for radiators that are long and narrow. These antennas are used to create fan beams, for example. Much like the planar scanner, these systems have also improved over the years because of the developments in planar scanners as well as through the development of high quality rotational hardware. Note that this scanner typically makes a linear scan followed by a rotation of the AUT or vice versa. In that regard, a planar scanner can be converted to a cylindrical one by adding an azimuth positioner to provide the rotation of the AUT.

The spherical scanner has increased in popularity because it can be used to investigate the volumetric pattern performance of low-gain radiators, such as cell phone antennas. In some situations, it is implemented by using two rotating systems to provide the spherical motion of the AUT, while the probe antenna is held stationary. In other cases, it is important that the AUT remains fixed. Thus, the probe antenna is rotated around the AUT.

In any case, all the probe data must be collected before the desired field pattern results can be obtained. In cases where the full spherical surface is scanned, the measured probe data can be used to provide the complete volumetric pattern. This is the only scanner system that can provide the full pattern, which makes it an important asset.

50.4 ABSORBER

The previous chamber dimensions did not include the absorber, because that is a major topic in its own right. To get a better understanding of the early absorber work, refer to Emerson's very good summary.¹⁶ Today, most engineers are familiar with microwave absorber in terms of its look-and-feel, but not as familiar with absorber performance limitations. As a result, this is the focus of this section.

Microwave, carbon-loaded, foam absorbers should be designed to have a low front face reflection and good attenuation as the incident electromagnetic energy propagates through the material and bounces off its back surface. To achieve these objectives, flat absorber material is typically constructed of multilayers. The first layer is very lightly-doped and does not attenuate the propagating wave very well, but it does have a descent front face reflection level. The next layers get more-and-more lossy such that the attenuation increases as the wave propagates through these layers. For this type of absorber, one typically measures the normal incidence reflection level at about -22 dB, which can be achieved with a traditional three layer design.

To control the front face reflection and still use lossy material, one can shape the front surface. The most popular shape has been pyramidal absorber. The front face reflection for this type of material is greatly reduced by the pyramidal shape. If the pyramidal absorber is properly designed and at least 2 wavelengths thick, it can achieve a -40 dB reflection level for normal incidence. Since wedge material is formed out of long 2D straight wedges and valleys as opposed to 3D pyramid tips, it does not transition as smoothly into free space. Therefore, it typically has a normal incidence reflection level that is between a flat and pyramidal absorber as one might expect. For a complete theoretical description of absorber scattering, refer to Yang et al.¹⁷⁻¹⁸

Normally, one will use a pyramidal absorber in antenna chambers, so let's focus on this type of material. Note that if pyramidal absorber is 2 wavelengths thick, it can achieve a -40 dB reflection level for normal incidence. Furthermore, the lowest reflection level achievable for a single pyramidal absorber panel is about -55 dB. It has been found through experimental studies that the lack of quality control in making the material leads to random variations within the material. These variations cause the -55 dB limit for a single panel. However, that does not mean that the total reflection level for a whole wall of absorber is limited to -55 dB. That is certainly not the case for a back or terminating wall in a chamber. It has been shown many times that the backwall scattering for a compact range chamber can be as low as -80 to -90 dB. This occurs because the absorber scattering comes from different random features within each absorber panel. That being the case, the scattering from individual absorber panels adds in power as opposed to adding in voltage. Therefore, the backwall scattering will drop to very low levels for a large backwall. Refer to DeWitt and Burnside¹⁹ for more details on this topic.

Before examining the angle of incidence performance, it is important to consider the low frequency performance of the absorber. Note that for normal incidence the reflection level degrades rather rapidly when the absorber is less than two wavelengths thick. This happens because the pyramidal shape is used to transition between free space and the absorber material media. For less than two wavelengths, this transition is too abrupt and the tips and valleys of the absorber are clearly seen in a time domain evaluation of the material. So at the low frequency end of the absorber operational bandwidth, the scattering appears to come from specific positions within the material. This means that one can place

the absorber tips and valleys into a Chebyshev height layout.²⁰ This height layout causes the various absorber tips and valleys to destructively interfere with each other.

Let's examine this approach using a one wavelength pyramidal absorber that is laid out in a Chebyshev pattern, which will reduce the reflection level by roughly 15 dB. So, a one wavelength pyramidal absorber in a Chebyshev pattern will have about a -50-dB reflection level. This is getting very close to the high frequency limit of -55 dB. As a result, the Chebyshev concept can be used to increase the low frequency bandwidth of pyramidal absorber. This will be especially important for low frequency chamber backwalls, which have traditionally used very thick pyramidal material. Taking another example, a 4' pyramid in a Chebyshev layout will result in a normal incidence reflection level of about $-30 - 15 = -45$ dB at 200 MHz. That means that it will work as well as a 12' pyramidal absorber panel, which is a huge reduction in cost and space. In addition, all the droop problems associated with using 12' pyramidal absorber can be avoided.

Absorber has been a mainstay in the design of chambers. As a result, the absorber companies have continually sold large volumes of this material. However, they have not provided new materials. That is a shame because antenna measurement requirements are continually getting more and more stringent and really require enhanced chamber designs. One area that has been very troubling is associated with the reflection performance level versus angle of incidence. Note that present-day materials degrade very rapidly after about 50 degrees. The only way to overcome this poor performance is to make the material thicker. Well, that solution is really good for the absorber companies but not for the antenna measurement community. Furthermore, one needs to really understand what is happening here.

As an incident field starts to approach grazing incidence, the material begins to look like a material media that is different than free space. Since a wavefront cannot propagate along this interface between the two media, it has to be totally reflected or nearly so. Thus, the basic physics causes this degraded performance, which means that some other material or design concept must be used to solve this situation. To make matters even worse, one has to measure the angle of incidence from the front face of the absorber because that face is where this undesired reflection occurs. Therefore if one increases the thickness of the absorber, the angle-of-incidence will reduce even further and the material will have to be made thicker and thicker. Therefore, traditional pyramidal absorber does not work well for wide angle incidence unless the undesired absorber scattering can be designed to miss the test zone and be absorbed in the backwall material.

Recently, there has been great interest in using the Chebyshev concept to help this angle-of-incidence problem. Since the absorber front face is causing this poor reflection level at lower frequencies, one can create a Chebyshev pattern for this situation. In this design, the area of concern is, for example, the sidewall scattering that is reflected into the test zone. Since this area is illuminated at say 60 degree, it will scatter a significant amount of energy into the test zone. If this sidewall material is then laid out in a Chebyshev pattern, it can be used to decrease this scattering by an additional 15 dB. That is very important in that it can greatly reduce the size of the chamber needed to perform a desired set of measurements. The details associated with the design of the Chebyshev absorber layouts can also be found in Gau et al.²⁰

If there is still too much sidewall scattering left because of this near grazing situation, consider a tapered R-Card fence approach, which is described in a later section.

Chamber Absorber Treatment

This section will describe how the absorber treatment should be determined for the various chambers. First, pyramidal absorber should be used in all regions of the chambers except for the sidewalls, ceiling, and floor of a tapered chamber. This special case will be examined later.

The first step is to determine the full operational frequency range for the chamber. Then the lowest frequency is used to determine the thickness of the absorber. For example,

suppose that the chamber operates down to 1GHz, in which case, the wavelength is roughly 12". Therefore, one should use $2 \times 12"$ or 24" pyramidal absorber, which is the proper choice for the front or transmit wall, sidewalls, ceiling, and floor. The back or receive wall should be lined with thicker material because it must completely terminate the very strong incident wavefront. This thicker material could be 36" standard pyramidal absorber or better yet, 24" pyramidal material laid out in a Chebyshev pattern.

Once the chamber is laid out with these materials, one needs to examine the angle-of-incidence question. Perhaps the sidewall, ceiling, and floor are too close to grazing. If that is the case, then one should see if the Chebyshev sidewall layout would provide the desired performance. If that is not good enough, then one has to use R-Card fences or increase the width and height of the chamber. In any event, make sure that all these issues are examined before the facility has been placed under contract.

Now, let's return to the tapered chamber absorber treatment. In this case, the source field basically propagates from the vertex of the tapered chamber to the test zone. That means that the field is propagating in a radial direction or at grazing incidence on the sidewall absorber. Since this field cannot propagate along the absorber boundary, the major issue is the excitation of the tapered chamber near the vertex. This excitation must be done in such a way that the source phase center is located at or near the vertex, which is surrounded by absorber. It has been done in at least two ways, one method has been to place a TEM horn at the vertex formed near the front face of the absorber. The TEM horn has straight walls that allow the source field to expand outward from the throat of the horn and the vertex of the tapered chamber. This TEM horn cannot be terminated abruptly into the absorber because that is too harsh a transition. To solve this problem, the TEM horn can be terminated in a tapered R-Card surface.⁴ Using this approach, the TEM horn metal walls slowly disappear and the presence of the absorber is transitioned in a much more gradual sense. Thus using this tapered R-Card TEM horn, the source field emanates from the vertex of the tapered chamber and propagates between the TEM metal walls that slowly disappear. This results in a very smooth transition of the source fields as they continue to propagate toward the test zone. So this causes the field within the tapered R-Card TEM horn to transition from a basic TEM horn mode to a fundamental horn mode with absorber walls.

The second antenna concept to successfully feed a tapered chamber has been a quad-ridge approach.⁵ This is a better approach because it is dual-polarized. With this method, the quad-ridge throat is placed near the vertex of the tapered chamber. The ridges are formed as extended elliptic ridges that are terminated into the absorber. For low frequencies, the ridges are not that long in terms of the wavelength so the absorber acts as a smooth termination. At higher frequencies, the energy naturally radiates off each curved shaped ridge as a normal creeping wave mechanism. As a result, the quad-ridge antenna is a very broadband radiator that can be used to properly excite a tapered chamber.

So from the previous discussion, one can see that the major issue is the proper excitation of the tapered chamber itself. Once the proper excitation is created, the field strength along the sidewalls is very weak and one can use much thinner material. Since this field is propagating along the down range direction of the chamber, one should use wedge material on the sidewalls, ceiling, and floor. Note that the wedge material is specifically designed for energy to propagate along its length provided that the edges of the wedge material are aligned with the direction of propagation. Near the throat of the tapered chamber and surrounding the feed antenna, one does not want to use wedge material because it is polarization sensitive. Thus, flat material should be integrated around the feed antenna. Naturally, this flat material must slowly transition into the wedge material. Note that the wedge material should extend from this transition region to the back or receive wall. The backwall should be lined with pyramidal material in a Chebyshev layout as done for the other chambers.

Before leaving this topic, there has been one other major issue associated with tapered chambers. Because the AUT can interact with the sidewalls, ceiling, and floor, many tapered chambers have been lined with pyramids laid out in a diamond shape and surrounding the

test zone. It is true that this material will reduce the interaction term between the AUT and surrounding walls, but it causes other problems that can be even more serious. Recall that the source radiation is propagating down the chamber and that the chamber becomes rectangular in the vicinity of the test zone region. This means that these pyramids surrounding the test zone will be illuminated at nearly grazing incidence. This is serious because each of these absorber tips will scatter energy into the test zone. This will create not only a stray signal but will significantly impact the polarization performance of the chamber. As a result, one needs a better solution to solve this AUT/wall interaction issue.

So what can be done to reduce the AUT/wall interaction term without causing these undesired stray signal affects? Recall that the wedge material was used on the sidewall in the horn section because it forms straight lines that align with the incident propagation direction. Therefore, it is very important that this straight-line flow of the absorber continues all the way to the backwall. To reduce the AUT/wall interaction term, one can again resort to the Chebyshev concept as described in Gau et al.²⁰ In this case, the edges of the wedges are laid out to form different heights. This causes the scattering for normal or nearly normal incidence to be greatly reduced in that the individual wedges tend to destructively interfere with each other. Thus, the wedge edges form continuous lines all the way down the chamber to the backwall and the AUT/wall interaction term is reduced by about 15 dB, which is about the difference in scattering levels for normal wedge and pyramidal material. Therefore, the Chebyshev wedge layout provides an ideal solution for this problem. In order to learn more about this tapered chamber layout concept, one should refer to Burnside et al.⁶

50.5 TAPERED R-CARD FENCES

As stated earlier, absorber does not work well at or near grazing incidence because the reflection coefficient approaches unity. In fact, normally, one should not use pyramidal material beyond about 50 degrees. In some situations, one can use pyramidal absorber in a Chebyshev layout to achieve reasonable performance out to about 60 degrees. But whatever the angular limit, one can clearly not use a traditionally available absorber at wider incidence angles. So again, what does one do for these wide angle situations? This is the focus of this section.

The traditional approach to solve grazing incidence and strong wide angle scattering has been to use an absorber fence. By using an absorber fence, the nearly grazing incidence signal on the original absorber is now nearly coming at normal incidence in terms of the absorber on the fence. Since it is nearly at normal incidence, the fence absorber works very well. So, the undesired grazing incidence scattering from the wall, floor, or sidewalls is removed by the fence. However this is basically a tradeoff of one problem for another. Let's assume that the fence absorbs all the incident energy. That means that the incident field that hits the fence is completely absorbed but the field that does not hit the fence is not absorbed. This creates a very harsh discontinuity in the incident field that is virtually the same as seen for a metal plate as discussed earlier. This undesired scattering is basically a diffraction that comes from the terminating edge of the fence. Since these fences are usually placed across the width of the chamber, the diffraction cone coming from the terminating edge of the fence will diffract a very strong signal into the test zone. Thus, the diffraction associated with the edge of the fence must be controlled. This can be done using a tapered R-Card fence.

A sample tapered R-Card fence is shown in Figure 50-5. This fence is formed out of a resistive material that varies from a perfect conductor at the bottom to a very large resistance at the top. This variation in the resistance means that the incident field will be perfectly reflected at the bottom of the fence; however, the reflection level will decrease as the resistance increases. At the very top of the R-Card fence, the resistance is usually on

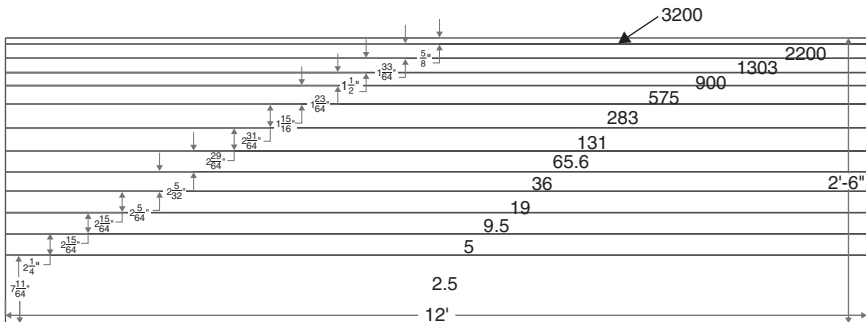


FIGURE 50-5 Tapered R-Card fence showing the resistance values for a sample design

the order of 3000 ohms per square. For such a large resistance level, the incident field is hardly impacted by this portion of the fence. Therefore, the incident signal transitions in a very smooth manner from being completely transmitted at the top of the fence to being completely reflected at the bottom. As a result, the R-Card fence is basically a fence that can block the grazing or nearly grazing signals on the absorber without diffracting. Thus, it has countless applications in chambers and outdoor ranges.

It has been shown by Gupta and Burnside²¹ that one can use these tapered R-Card fences to block the ground bounce in a ground bounce range. One might ask, why is that so significant? Well, it allows the set up of an outdoor measurement facility without being impacted by the ground bounce, which means that the measurement can be taken over extreme bandwidths. That cannot be done easily using a traditional ground bounce range unless the source height above the ground is varied versus frequency.

Before leaving this topic, there is one other modification of this concept that has been found to be very useful, especially for indoor applications. Recall that an R-Card fence is basically a reflecting surface that does not diffract. Well for most applications, one would really like an absorber fence that does not diffract. This can be done as well. The tapered R-Card fence is basically constructed of resistive material. It has been known for many years that one can take resistive sheets and form an absorber, such as a Salisbury screen. Using this idea, one can construct a five layer design such as shown by Lee and Burnside.²² This design provides at least a -22 -dB reflection level for normal incidence from 2–18 GHz. So this absorber consists of a bottom metal layer and then four resistive layers separated by free space or foam.

So now, it is understood how one creates an absorber out of these resistive sheets. To make the absorptive R-Card fence that absorbs but does not diffract, one has to take each of the resistive layers and taper the resistivity going from the original resistance level to a very high resistivity, say 3000 ohms per square. Such a design is also shown in Lee and Burnside.²² Note that the fence is about 5 wavelengths tall at the lowest frequency and can be used very effectively to remove the wide angle scattering associated with traditional absorber. This concept will be shown in the next section to be extremely useful for hardware-in-the-loop chambers.

50.6 HARDWARE-IN-THE-LOOP CHAMBERS

Hardware-in-the-loop chambers are widely used to examine the performance of real operational hardware when it is illuminated by a wide variety of radiators. In these chambers, the real hardware is inserted into the test zone and illuminated by various sources mounted on a transmit wall. These are typically rectangular chambers with fixed antennas mounted

at various locations on the transmit wall or moving antennas that can be used to simulate various real-world scenarios. The antennas on the transmit wall typically come very close to the sidewall; thus, they potentially illuminate the sidewall absorber at or near grazing incidence. In evaluating a wide variety of these chambers, it has been found that one should use a rectangular chamber with absorber to handle the antenna locations near the center of the transmit wall. If an antenna comes close to the sidewall and the angle-of-incidence on the sidewall absorber is greater than 50 degree, then the Chebyshev layout can be used. If the angle-of-incidence is greater than 60 degrees, then R-Card fences are needed. The placement of the fences is typically located at the specular reflection position assuming that the reflection comes from the tips of the absorber. Since there can be many antenna positions that are close to sidewalls, one may need more than one R-Card fence. A typical sidewall design for a hardware-in-the-loop chamber is shown in Figure 50-6. Note that the sidewall absorber is laid out in a Chebyshev pattern and that absorptive R-Card fences are triangular in shape and positioned to remove the specular reflections for each of the antennas close to the sidewalls. The same treatment should be done on the floor and ceiling if the antennas come close to those surfaces as well.

Note that the tapered R-Card fences can be used to hide the junction associated with absorber being thinner closer to the source antennas and thicker farther away. Also, the fences become larger away from the source antennas because they need to block more energy coming from the antennas farther from the sidewalls.

Besides the chamber issues, hardware-in-the-loop testing involves using real-world systems that are not meant to be mounted in a chamber. Furthermore, to simulate the real-world operation of these systems, they must appear to be moving at high speeds and placed in very demanding situations. That being the case, the motion control hardware must be massive and yet not impact the measurement. This can only be done to a certain degree because of the size of the equipment needed to properly hold and move the hardware being tested. First, the mounting structure has to be designed to keep the motion control system as far as possible from the AUT. Because the hardware must remain in the test zone, it must be close to the center of rotation. To accomplish all these desires at the same time, one should use a mount structure that places the very large azimuth rotator as close to the floor as possible. The azimuth rotator then holds a tilted boom arm that extends vertically into the region just

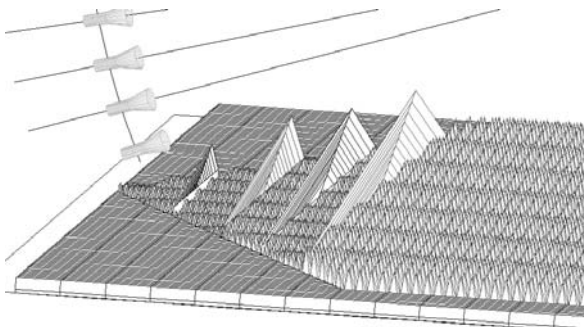


FIGURE 50-6 Typical hardware-in-the-loop chamber showing the R-Card fences on the sidewall along with the pyramidal absorber laid out in a Chebyshev pattern

behind the test zone. This tilted arm holds the elevation rotational hardware. The tilt in this arm causes the incident signal to be reflected downward away of the test zone, much like a pylon that is used for RCS testing. In fact, this tilted arm could be designed to have an ogival cross-sectional shape that is pointed in the front and back. Using this approach, one minimizes the impact of the very large azimuth rotator and associated vertical arm that is normally used to hold the real-world hardware in the center of the test zone.

The last item that must be hidden is the very large elevation rotation hardware that is mounted directly behind the radiating system. This is done by attempting to model the actual hardware mounting structure. Note that this mounting hardware is probably metal, or least conducting. So a conducting shroud should be placed between the radiating system and the elevation rotational equipment. It will hide the elevation rotational hardware but will have an exposed trailing edge. The potential edge diffraction from this edge could also impact the desired measurement. This diffraction can again be minimized by using a tapered R-Card concept for the trailing edge of the shroud. Using all these techniques, the rotational hardware is pretty well hidden and the measured data should be as realistic as possible considering the size of the rotational hardware needed to safely maneuver real-world hardware during these hardware-in-the-loop tests.

50.7 GAIN, EFFICIENCY, AND BORESIGHT MEASUREMENTS

For small antennas, the efficiency is the major concern.^{23–24} On the other hand, for most antennas, the gain is typically the parameter that must be measured. This is normally done by comparing the gain of the unknown antenna with the gain of a known antenna. The first step in this process is to find the boresight direction, which is typically determined by measuring the main beam pattern at several selected levels and then assuming that the true boresight direction is half-way between the angular locations of a given selected level. The use of several pattern levels provides averaging to refine the boresight location. Once the boresight direction and gain are known for the AUT, one then measures the boresight direction and gain for a standard gain antenna using the exact same measurement hardware. From these two measurements, one can then determine the difference in gain between the known and unknown antennas. This difference is then added to the known gain to determine the gain of the unknown antenna. This seems rather simple, but it has many pitfalls.

First, it must be assumed that the field quality is perfect across the whole test zone used to measure the two antennas. That is typically not possible for a high gain antenna because it is illuminated by a field that varies somewhat in amplitude and phase. For a compact range, this could be as much as a dB or more in some larger applications. Next, there is the question of polarization. The AUT may be circularly polarized, but the known antenna is linearly polarized. All of these potential issues must be resolved before one can accurately determine the gain of an unknown antenna.

In recent years, there has been a great interest in knowing the gain to within tenths of a dB. That means that one must account for every possible error term in their gain measurement. The measurement uncertainty is assessed through an error budget projection comprised of the arithmetic sum of the mean error components and an RSS sum of the standard deviations. As a result, it is very difficult to verify that the desired gain level has actually been achieved. To overcome this situation, it is very useful to have different ways to make these gain measurements. If the desired level of performance is achieved, one should be able to verify the results using dramatically different gain tests.

Let's consider that it is our desire to measure the gain through an RCS measurement in that it is well-known that the antenna-mode backscatter term is directly proportional to the

gain of the antenna squared.²⁵ Since the RCS level can be very accurately calibrated, one does not need to use a standard gain antenna. Therefore, this method removes the potential errors associated with a second antenna measurement.

That being the case, how does one measure the antenna-mode RCS scattering level of a high gain antenna. First, one needs to understand that the total scattering is made up of two terms: 1) the antenna-mode scattering and 2) the structure scattering. These two terms can be separated by matching the antenna port. In this case, the antenna-mode signal is fully absorbed in the matched load and the total scattering comes from the structural scattering. So this offers one way to isolate the antenna-mode scattering from the structural scattering. But this approach is not very useful in practice, because it is often times very difficult to find a perfectly-matched condition. That is especially true for broadband applications. So, it is strongly suggested that a time domain method be used, which is often the preferred measurement approach for high-quality RCS measurements.

Using the time domain, it's possible to separate out the various time responses, because the structural scattering comes at certain times and the antenna-mode scattering at others. In fact, the antenna-mode scattering can be moved in time by changing cable lengths. This is done by attaching a length of cable to the antenna port. At the other end of the cable, it is connected to a multi-throw switch that has its output ports connected to a set of different length cables terminated in short circuits. The RCS of the antenna is then measured with the antenna in the boresight direction. This measurement is performed over the available bandwidth of the antenna. The same measurement is then repeated with the multi-throw switch connecting different lengths of cable. This measurement is repeated a number of times so that one can clearly separate the structural scattering term from the antenna-mode term. Once the antenna-mode term is isolated, it can be calibrated by a flat plate of the same aperture size as the AUT. This means that both apertures are exposed to the same exact field. This will provide the desired antenna-mode RCS level. Once the cables connected to the antenna port are calibrated using very accurate network analyzer techniques, the only unknown is the AUT gain. This RCS gain measurement technique is described in Rudduck et al.²⁶

This RCS method is very useful for high-gain antennas. In addition, the large plate that is used for the RCS calibration can also be used to boresight the high-gain antenna. This is done by using the mechanical alignment of the plate relative to some feature on the mounting structure that will ultimately hold the antenna. So, this plate is used as a mounting standard for boresighting the antenna in a mechanical sense. The high-gain antenna is then boresighted so that its beam is fully aligned with the flat plate peak scattering. This is done by appropriately shimming the antenna to get the antenna radiated main beam to align with the flat plate RCS main beam. Note that the plate can be mounted some distance behind the high-gain antenna so that an absorber can be added to make sure that it does not interfere with the forward radiation from the antenna. Thus, the flat plate serves two functions. It is used to very accurately align the boresight direction of the main beam with the mounting structure and to also provide a very accurate RCS calibration standard. Since the potential errors in this method are dramatically different than those in the substitution method, it is an excellent way to either measure the gain or verify the substitution method results. Finally, one should realize that the RCS calibration and boresight plate should be removed before any full patterns are attempted, especially in cases where the backlobe pattern region is of interest.

This RCS method is not very useful for medium or low-gain antennas. Thus, another method is needed for these cases. One can use an S11 measurement technique.²⁷ Using this approach, one takes the AUT and mounts it in front of a very large polished flat plate. The flat plate reflection then re-illuminates the AUT. This illumination comes from the perfect image of the AUT. In other words, one is measuring the S12 transmission between two identical antennas. Since the S11 port can be very accurately calibrated and this S12 coupling term can be separated in time, one can obtain the S11 measured results that are

fully calibrated and the gain is the only unknown provided that the antenna's phase center is accurately determined using standard pattern measurement techniques. Consequently, this method again does not need a standard gain antenna, so it is not susceptible to errors related to the substitution method. Thus, it is a wonderful method to obtain gain data quickly and accurately for low or medium gain antennas.

Traditionally, gain measurements have focused on the magnitude alone. For most applications, that is all that is needed. However, there are new broadband applications that are becoming more important all the time. That being the case, one is also very interested in the main beam phase behavior versus frequency. Recall that for broadband systems, the magnitude and phase variations can cause undesired modulation responses. Thus, the phase measurements cannot be ignored. There are many standard gain antennas that one can use to perform normal gain magnitude measurements; however, gain phase tests for standard gain horns have not been done to any great degree. Thus, new methods must be developed to perform these broadband phase tests.

Since neither the RCS nor flat plate reflection tests require a secondary gain standard, one can use either of these tests to collect both the magnitude and phase. Let's consider the flat plate test. In this test, the S_{11} response associated with the flat plate reflection term is isolated in the time domain. From this data, one knows both the magnitude and phase data. The phase data is related to the path length between the horn and its image, plus the phase variation associated with AUT. The path length between the AUT and its image can be determined. This leaves the AUT phase as the only unknown. Thus, this method can be used for future broadband phase tests as well. Note that similar concepts can be used to extract the AUT phase variation using the RCS test.

50.8 POLARIZATION MEASUREMENTS

In order to take any pattern measurement, one must also be concerned with the polarization of the AUT. To fully understand the polarization properties of the AUT, it has been found most useful to measure both linear polarization patterns because it is generally easier to achieve high polarization purity with a linear state. If these patterns are taken so that they are phase coherent, then one can combine these two polarization results in any way that is needed to fully characterize the AUT's polarization performance. The difficulty in measuring these two patterns is to insure that they remain stable in magnitude and coherent in phase. To properly insure the stability and coherency, one can take advantage of the rotational hardware used to hold the AUT during these measurements. This will be discussed in more detail later.

As shown by Ludwig,²⁸ the third polarization definition is most useful and found to be consistent with how one should take the measurement in terms of keeping the AUT in the middle of the test zone. Let's suppose that a measurement facility has a roll-over-elevation-over-azimuth motion control system. Using this hardware, there should be a tall vertical pole between the elevation-over-azimuth rotator, which is mounted on the floor. This pole should then be oriented perpendicular to the direction of plane wave propagation within the chamber. Then, the elevation rotator should be used to boresight the AUT. This boresight procedure is done using the azimuth and elevation rotators. Once the antenna is boresighted the elevation rotator angular position is fixed for all subsequent measurements. Now, the polarization or roll axis is adjusted to align with a desired linear polarization state and the pattern is taken by rotating the azimuth positioner from 0 to 180 degrees in step sizes that satisfy the sampling theorem. Once this azimuth pattern is collected, the azimuth axis should return to 0 degrees going in the opposite direction to the pattern motion. Next, the

polarization axis is rotated to a desired step size and the azimuth pattern is taken again. This process is continued until the full volumetric pattern is obtained. Recall that this set of patterns is taken for one linear polarization. So, the source antenna should then be rotated by 90 degrees to obtain the orthogonal polarization pattern. Thus, the full volumetric pattern is repeated for this new source antenna polarization. This approach will provide the complete pattern description for the AUT provided as stated earlier that the two polarization pattern sets are stable in magnitude and coherent in phase.

To insure that the two polarization patterns are stable and coherent, one simply needs to recall the data collection process. First say that the source polarization is horizontal and the AUT is also horizontally polarized for the initial roll axis angle of 0 degrees. That being the case, the source and AUT polarizations are aligned at 0 degrees. Now consider the second polarization measurement, the source polarization is now vertically oriented and at a roll axis position of 90 degrees, the AUT polarization is also vertical. That means that the same boresight measurement is being repeated using either the horizontal and vertical source polarizations. Therefore, these two measurements can be used to insure that the two polarization patterns are stable in magnitude and coherent in phase.

Before leaving this topic, it's important to consider one more aspect associated with this measurement approach. The concern here is that the large AUT remains in the test zone during the full measurement. As shown earlier, the test zone size is a very important parameter in the design of a measurement system, in that the cost of the measurement facility tends to be directly correlated to the test zone size. As a result, it is very important to keep the test zone size as small as possible. That being the case, one should also want to make sure that the AUT remains centered in the test zone; otherwise, the AUT will go outside the test zone resulting in undesired measurement errors. With this in mind, the roll-over-azimuth approach is ideal because the AUT can be mounted on the support structure to insure that it is always located inside the test zone boundaries.

50.9 ANTENNA TEMPERATURE MEASUREMENTS

Normally, receiver noise dominates over antenna generated or received external noise. However for low-noise systems or systems that tend to look toward the sun or other hot spots, one must understand how to take proper antenna temperature measurements. Normally, these measurements are not taken indoors because it involves having the AUT illuminated by the appropriate real-world external noise sources, whereas a chamber has an ambient 290K background. Even so, one can still gain great insight into antenna temperature measurements by properly characterizing the antenna properties. First, one needs to fully measure the volumetric pattern of the AUT. Then each of the losses in the antenna system must also be identified. In most cases, it is much easier to measure these parameters indoors. Of course, there are measurements that can be done outside using known temperature sources. Once these parameters are fully understood, one can then create various scenarios to investigate the impact of certain real-world situations and the impact on system performance.²⁹⁻³⁰

One parameter that is very important to understand is the volumetric pattern. This must be known to very low levels because any real-world system can be illuminated by very strong noise sources. If these strong sources illuminate the AUT in a direction with a high sidelobe, it can cause significant noise and the system performance will degrade. Thus, the sidelobe performance can be a very serious matter and require measurements that are accurate even for huge dynamic ranges. That being the case, one has to be very concerned about stray signals in a chamber at wide angles.

50.10 CHAMBER EVALUATION

A major problem with antenna/radar test ranges today is the lack of knowledge of their test zone fields. Ideally, the test zone fields should be a uniform wavefront that represents the desired illumination of the AUT. In practice, depending on the range configuration, the test zone fields will show variations in magnitude as well as phase. Some of these variations are associated with the range geometry (spherical wavefront, taper in compact ranges, and so on), whereas some variations are caused by undesired (stray) signals present in the range. The sources of these stray signals tend to be localized, for example, a bad piece of absorber, feed spill over in compact ranges, and so on. The stray signals affect the quality of the measured data very differently than the variations in the test zone fields due to the range geometry. Let's say, a given measurement facility has a dB taper in the test zone fields due to the range geometry. This taper variation will lead to some drop in the measured sidelobe levels as well as some small reduction in the measured gain. On the other hand, if the test zone fields have a ripple due to the presence of stray signals, it can lead to large errors in the measured sidelobe levels. For example, suppose that the stray signal appears at 20 degrees with respect to the desired plane wave term and is 30 dB below the desired illumination. Furthermore, the AUT has a sidelobe level of -40 dB at 20 degrees. As the AUT is rotated toward the 20 degree direction, the desired signal level should be -40 dB; however, the main beam now points directly at the -30 dB stray signal and the received signal is -30 dB. Thus, this stray signal can cause a 10 dB measurement error. To reduce the measurement errors, stray signals must be found in a test zone and removed if at all possible. If the stray signals cannot be removed, they should be avoided during the measurement. Finally, if avoiding or controlling the stray signals is impossible, then their impact on a given measurement must be understood. This understanding will be explained at the end of this chapter. It is a new concept that can be used to answer the age old question: "Can your facility be used to accurately measure my antenna?"

A further advantage to chamber evaluation is an understanding of those factors that limit the facility performance. Generally, several factors contribute to different degrees to the measurement uncertainty and reducing each of the individual factors that limit performance is accompanied by different economic impacts. By using the chamber evaluation to determine the limiting factors and addressing the scope of ways to deal with the individual factors, the cost/benefit analysis for facility upgrades can be understood.

The first step in the range evaluation process is the determination of the stray signals in the chamber. This is best done in an antenna measurement facility through field probing. A *prober* should be installed in the center of the test zone. It can be an x/y or ρ/ϕ *prober system* that is oriented to scan across the full width and height of the test zone. The probe data is collected by moving the probe antenna in steps that satisfy the sampling criteria. At each step position, the probe data is collected over as wide a bandwidth as possible. This measured probe data contains the desired signal plus any undesired stray signals that are seen by the probe antenna. Note that the probe antenna should have as wide a pattern as possible so that the full half space forward of the prober is properly evaluated. Since one only wants to collect stray signal data in the front of the chamber, the probe antenna should be mounted on an absorber panel so that it does not see stray signals behind the probe antenna, including the scattering from the probe structure itself. Also, since most chambers have very broadband frequency capability, the probing should be done to cover the full operational bandwidth.

Once the test zone of a measurement facility has been probed, the probe data can be displayed in various formats. The most common format is a plot of the amplitude and phase as a function of probe position at a fixed frequency. Almost everyone has seen the probe data in this form. Typically as one looks at this type of data, it's possible to notice that it is

made up of small complex amplitude and phase variations added to the desired illumination term. These variations are not simple because they represent many small stray signals that emanate from various places within the chamber.

In order to make sense out of these complex responses, one must resort to some post-processing to resolve the various signals seen in the test zone. Many techniques for post-processing of the probe data have been reported in the literature. Gupta³¹ provides a summary of some simple yet very useful techniques for post-processing of the probe data such that the sources of the various stray signals in the antenna/RCS measurement facility can be properly located. These techniques include near field focusing, direction of arrival (DOA) estimation, time of arrival (TOA) estimation, time and direction of arrival estimation (TADOA), and so on.

TADOA estimation³² is a very useful approach for locating the sources of stray signals in an antenna/RCS test facility. In this approach, the probe data at each frequency is first transformed to the angle domain and then the frequency domain data is transformed to the time domain. Thus, each stray signal in a TADOA plot is isolated in terms of its specific time delay and angle-of-arrival. For example, Figure 50-7 shows the TADOA spectra of a test range that was evaluated from 250 MHz to 750 MHz. The main lobe in the middle of the plot corresponds to the desired illumination, whereas the other peaks correspond to stray signals. One can clearly see a strong stray signal at -50 degrees. This stray signal arrives 24 nanoseconds before the direct illumination. With this information, one can determine the true stray signal path within the chamber by tracing back from the center of the test zone along the angle-of-arrival direction. Once the true path is found, it should have the correct relative time of arrival versus the desired term as well as the proper direction-of-arrival angle. If one thinks about this approach in some detail, it becomes clear that each undesired stray signal path can be identified. Once the true stray signal sources are determined, one should immediately attempt to correct these undesired error terms. Note that the probe data must be re-collected after the chamber modifications to insure that the error term or terms have indeed been corrected. The beauty of this approach is that one can clearly isolate each stray signal and be assured that it has been removed even though the chamber may have many other small stray signal terms. Recall that if one corrects one small stray

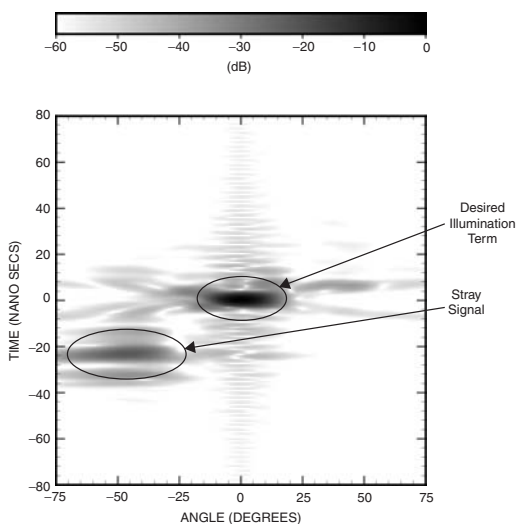


FIGURE 50-7 TADOA spectra of a test range from 250 to 750 MHz

signal out of many, very little difference in the raw probe data will be seen, but the chamber improvements will be very clearly visible in the plots of TADOA spectra.

Now let's suppose that everything has been done to minimize the stray signal error terms in a given chamber. This means that one must live with what remains. That being the case, how can it be determined whether the chamber can be used to measure a given antenna? To answer this question, let's suppose that the chamber has been fully probed at the desired measurement frequencies. That means that the field incident on the AUT is known through the probe data. This measured probe data could then be used to create a plane wave spectrum that will ultimately illuminate the AUT at each desired frequency. Since the AUT was designed using some type of numerical simulation code, it has probably been fully characterized in terms of its amplitude and phase response versus various angles of incidence. In many cases, the full volumetric pattern has been computed. If the full pattern is not known, at least the dominant regions of the pattern are known. When this AUT is mounted in the test zone, it will be exposed to the same illumination as seen with the probe data. Thus, the plane wave spectrum of the measured probe data is a very good representation of the actual AUT illumination. That being the case, one can take this incident plane wave spectrum and illuminate the AUT in terms of a full simulation. Note that now the simulation is being used to determine what the measured pattern data should look like based on the measured incident fields in the test zone as opposed to the desired plane wave illumination. This will result in what one could call a "simulated measured pattern," or SMP for short. This SMP can then be compared with the predicted pattern using the desired illumination. This comparison can be done in many ways. But the bottom line is that one can use this technique to determine whether or not a given measurement facility can be used to measure a specific antenna. In fact, the expected error bounds could be known before any measurements are even taken. Thus, one can answer the age old question: "Yes, my chamber can be used to measure your antenna and here is the SMP data to prove it!"

One could become very uneasy about this chamber evaluation process because it could expose stray signals that cannot be corrected, which might result in the loss of potential chamber support. That is not necessarily the case, because the existence of a stray signal in a chamber is not really the problem. The problem is that the stray signal can impact an important measurement, which is the bottom line. To illustrate this point, let's again consider the feed spillover error term in a compact range. Since the feed antenna in a compact range is normally mounted near the floor, this feed spillover error term comes from a down look angle as seen from the center of the test zone. Let's suppose that the look angle for this error term is 20 degrees down and a large reflector antenna system is being measured in the test range. In this measurement, the full volumetric pattern data is needed. If one takes this pattern data by allowing the main beam to fully scan around the chamber, it will eventually scan down 20 degrees and measure the feed spillover error term through the main beam. The pattern measurement for this case is very likely to be in error. In fact, this measurement error could be tens of dBs, as described several times earlier. On the other hand, if the chamber operator is well aware of this error term and its impact on similar measurements, he or she would know to take the full volumetric pattern in such a way that the main beam never scans down. This way, the feed spillover is still present but it does not affect the measured results nearly as much as it would have otherwise. So, the value of this type of knowledge is priceless.

The previous example was very useful because it illustrates how one can avoid undesirable stray signals. This is well and good but how does one use this information for more complex measurements that have multiple beams or fan beams or shaped beams, and so on? Can one still find ways to avoid the stray signal errors for these very complex pattern measurements? Yes, this can be done through the simulation method described earlier.

Let's suppose that the SMP is computed and compared with the desired pattern. It is then noticed that the difference between the SMP and desired pattern is greater than what

is acceptable. This could very well be the case. But all is not lost at this point in that one now knows the error angles and some predicted error bounds for that specific measurement. Once this is known, one can take the pattern measurement in a different way and obtain data that is correct in the direction or directions that failed before; however, it fails in some other directions. That is not a problem, because one can anticipate these errors and their associated predicted error bounds. If the actual measurements follow these predicted findings, it is perfectly acceptable to use the correct data where it was predicted to be correct through simulation. However this will only work if the measurement and simulation experts work together to achieve their common goal of making accurate antenna pattern measurements.

The previous approach also solves another major issue with associated measurements in general. The antenna community has been divided into two groups: 1) the simulation group and 2) the measurement group. This has led to many confrontations between these two groups because the measured and calculated results oftentimes do not agree to the precision required in the evaluation of modern systems. That being the case, the two groups tend to defend their own turf and tend to point out the potential shortcomings of the other group. That is not right, in that both groups really have their own limitations. So they need to work together. This can be done using the approach presented here. In fact, the simulation group can determine the measurement error bounds before the measurements are even performed. Simulations can be used to determine if a given facility can be used to measure a given antenna with the desired accuracy. Unfortunately that is not happening today. Furthermore if the simulation group agrees that a given facility can do the measurement with certain restrictions, they cannot complain that the measurement is wrong in that their simulation fully justified the measurements. So, each group has a way to protect their own turf using this concept.

50.11 CLUTTERED ENVIRONMENT MEASUREMENTS

There are times that pattern measurements are taken and the results are not satisfactory. This can be the result of undesired stray signals that were never corrected. At this point, these error terms may not even be understood, yet accurate results are needed from this suspect facility. So, what can be done? This is the focus of this section.

To illustrate the value of this technique, let's suppose that the desired measurement is of a high gain antenna and the pattern is taken using a roll-over-azimuth rotational system so that the main beam is always scanned around the horizontal plane in the chamber. This being the case, the concern is associated with stray signals in the horizontal plane or nearly in the horizontal plane. To begin the process, the antenna is mounted in the center of the test zone and the first pattern is measured. Now let's examine a second pattern that can be taken to correct the errors in the first pattern. To grasp this concept, one needs to think about how a stray signal corrupts a given measurement. As stated earlier, it happens when the AUT main beam illuminates the stray signal direction and amplifies the stray signal level. Since a low sidelobe is being measured rather than the main beam, one needs a method that will naturally attenuate the main beam gain level without impacting the desired low sidelobe measurement. This can be done by taking a second pattern such that the two patterns add in-phase in the desired plane wave illumination direction and out-of-phase in the AUT main beam direction. It's possible to think of the summation of these two patterns as an array pattern that has a peak value in the plane wave illumination direction and null in the AUT main beam direction. This array pattern can be obtained if the second pattern is taken with the antenna positioned along a cross-range direction so that the plane wave illumination is always in-phase.

The position along the cross-range direction is selected so that the array pattern has a null in the AUT main beam direction for a particular azimuth pattern angle. Since this position changes with the pattern angle, the second pattern must be taken with the cross-range position changing with pattern angle. Yes, the motion control system has gotten more complex, but these two patterns can then be added together and the stray signal errors reduced by 20 dB or so. Furthermore, the actual measurement errors can be seen if these same two patterns are subtracted. If this subtracted pattern is negligible, the chamber measurements were correct because there were no undesired peaks in the subtracted pattern. If the subtracted pattern is relatively significant compared to the desired pattern, then the original measurement was in error and the actual error level can be clearly seen. This is a very powerful way to take measurements in a cluttered environment and still get valuable measured data that one can fully justify as being correct. This discussion has been greatly simplified in that it was simply used to highlight the method. For the full details, one should refer to Burnside and Gupta.³³

50.12 SUMMARY AND CONCLUSIONS

This chapter has attempted to examine indoor antenna measurements. It has been shown that modern measurement hardware has been so refined in recent years that it's now possible to have tremendous capability at your finger tips. Because of these awesome tools, S11 measurements can be taken with tremendous precision. For pattern, antenna temperature, and hardware-in-loop measurements, the major concern is stray signals that can contaminate the measured results. Therefore, one must be able to fully understand and control these undesired error terms as shown here.

Various indoor facilities have been examined in terms of their usefulness and limitations. The limitations can be specifically identified by taking probe measurements that characterize the test zone fields. By properly processing the probe data, the stray signal sources in any chamber can be isolated, identified, and potentially corrected. If they cannot be corrected, they should be understood so that any subsequent measurement can avoid these uncorrected stray signals.

In some cases, it is virtually impossible to know the impact of the residual stray signals on a complex antenna pattern measurement. That being the case, it has been proposed here that the antenna simulation team, which wishes to use a given chamber to measure their antenna, should take the measured probe data and create an incident plane wave spectrum. This plane wave spectrum can then be used through simulation to illuminate their antenna. The resulting pattern from these simulations can then be compared with the desired pattern. If they feel that the resulting simulated measured pattern is satisfactory, they will know that the given chamber will provide the desired results. This way, the simulation team will be able to justify using or not using a given facility for their measurements. In addition, the simulation team can propose back to the measurement team what could be done in terms of the measurement procedure within a given chamber to avoid certain errors that are critical to their specific antenna measurement. In any case, everyone knows what to expect so that the major conflicts that exist today between the simulation and measurement groups can be avoided to a large degree.

Finally, it has been shown that one can take measurements in a cluttered chamber provided that a second measurement can be obtained with the AUT moved in a cross-range direction while being rotated in angle. This second measurement is used to remove the stray signal errors. If these two patterns are added together, they form an array pattern that does not impact the desired plane wave illumination but places a null in the AUT main beam direction. Note that the null in the main beam direction will reduce the stray signal

error by about 20 dB. Since this is all based on measurements, it is a wonderful way to show the measurement performance of a given facility. On the down side, the motion control system becomes more complex because the AUT has to be moved in a cross-range direction as it rotates in angle. Even so, additional motion control systems can be an inexpensive fix as compared to modifying a very expensive measurement facility. In any event, it is important that one understands their facility and the impact that their facility has on any proposed measurement.

The bottom line here is that one can now use the tools presented in this chapter to answer the age-old question: "Yes, my facility can measure your antenna and here is the proof." Thus, it is hoped that the antenna community will eventually adapt these methods as standard practice.

REFERENCES

1. R. B. Dybdal, "Methodology to Project Antenna Measurement Accuracy," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 1989.
2. Antenna Standards Committee, "IEEE Standard Test Procedures for Antennas," The Institute of Electrical and Electronics Engineers, Inc., 1979.
3. T-H Lee, W. D. Burnside, "A Diagnostic Technique Used to Obtain Cross-Range Radiation Centers from Antenna Patterns," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 10 (October 1989).
4. W. D. Burnside, L-C. T. Chang, M. Gilreath, B. Smith, "An Ultra-Wide Bandwidth Tapered Chamber Feed," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 1996.
5. K-H. Lee, C-C. Chen, R. Lee, "UWB Dual Linear Polarized Feed Design for Tapered Chamber," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 2003.
6. W. D. Burnside, S. J. Essman, W. H. Theunissen, D. Steinberger, H. H. Chung, "An Enhanced Tapered Chamber Design," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 1999.
7. R. C. Johnson, H. A. Ecker, R. A. Moore, "Compact Range Techniques and Measurements," *IEEE Transactions on Antennas and Propagation*, vol. AP-17, no. 5 (September 1969).
8. W. D. Burnside, M. C. Gilreath, B. M. Kent, G. L. Clerici, "Curved Edge Modification of Compact Range Reflector," *IEEE Transactions on Antennas and Propagation*, vol. AP-35, no. 2 (February 1987).
9. I. J. Gupta, K. P. Ericksen, W. D. Burnside, "A Method to Design Blended Rolled Edges for Compact Range Reflectors," *IEEE Transactions on Antennas and Propagation*, vol. 38, no. 6 (June 1990).
10. T-H. Lee, W. D. Burnside, "Performance Trade-off Between Serrated Edge and Blended Rolled Edge Compact Range Reflectors," *IEEE Transactions on Antennas and Propagation*, vol. 44, no.1 (January 1990).
11. T-H. Lee, W. D. Burnside, "Compact Range Reflector Edge Treatment Impact on Antenna and Scattering Measurements," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 1 (January 1997).
12. A. D. Yaghjian, "An Overview of Near-Field Antenna Measurements," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 1 (January 1986).
13. D. Hess, "Readily Made Comparison Among the Three Near-Field Measurement Geometries Using a Composite Near-Field Range," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 2003.
14. A. C. Newell, "Error Analysis Techniques for Planar Near-Field Measurements," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 6 (June 1988).
15. R. C. Baird, A. C. Newell, C. F. Stubenrauch, "A Brief History of Near-Field Measurements of Antennas at the National Bureau of Standards," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 6 (June 1988).

16. W. H. Emerson, "Electromagnetic Wave Absorbers and Anechoic Chambers Through the Years," *IEEE Transactions on Antennas and Propagation*, vol. 21, no. 4 (July 1973).
17. C-F. Yang, W. D. Burnside, R. C. Rudduck, "A Periodic Moment Method Solution for TM Scattering from Lossy Dielectric Bodies with Application to Wedge Absorber," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 6 (June 1992).
18. C-F. Yang, W. D. Burnside, R. C. Rudduck, "A Doubly Periodic Moment Method Solution for the Analysis and Design of an Absorber Covered Wall," *IEEE Transactions on Antennas and Propagation*, vol. 41, no. 5 (May 1993).
19. B. T. DeWitt, W. D. Burnside, "Electromagnetic Scattering by Pyramidal and Wedge Absorber," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 7 (July 1988).
20. J-R. J. Gau, W. D. Burnside, M. C. Gilreath, "Chebyshev Multilevel Absorber Design Concept," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 8 (August 1997).
21. I. J. Gupta, W. D. Burnside, "Performance of an Experimental Outdoor RCS Range with R-Card Fences," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 2001.
22. T-H. Lee, W. D. Burnside, "Applications of Multilayer Resistive Strips (R-Card) in EM Measurements," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 2003.
23. D. M. Pozar, B. Kaufman, "Comparison of Three Methods for the Measurement of Printed Antenna Efficiency," *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 1 (January 1988).
24. R. H. Johnston, J. G. McRory, "An Improved Small Antenna Radiation-Efficiency Measurement Method," *IEEE Antennas and Propagation Magazine*, vol. 40 (October 1998): 40-48.
25. J. Appel-Hansen, "Accurate Determination of Gain and Radiation Patterns by Radar Cross-Section Measurements," *IEEE Transactions on Antennas and Propagation*, vol. 27, no. 5 (September 1979).
26. R. C. Rudduck, K. M. Lambert, T-H. Lee, W. D. Burnside, "Gain Measurement of Microwave Antennas Using a Backscatter Technique," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 1988.
27. I. J. Gupta, E. H. Newman, C-C. Chen, W. D. Burnside, "A Novel Method for Antenna Gain and Phase Calibration," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 2005.
28. A. C. Ludwig, "The Definition of Cross Polarization," *IEEE Transactions on Antennas and Propagation* (January 1973).
29. K. M. Lambert, R. C. Rudduck, "Measurements for the Verification of Antenna Temperature Calculations for Reflector Antennas," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 1992.
30. D. T. Copeland, R. B. Wendt, D. A. Myers, "Prediction of Equivalent System Temperature in the Outdoor Environment Using Indoor Measurement," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, 1998.
31. I. J. Gupta, "Stray Signal Source Location in Far-Field Antenna/RCS Ranges," *IEEE Antennas and Propagation Magazine*, vol. 46 (June 2004): 20-29.
32. I. J. Gupta, W. D. Burnside, E. K. Walton, "Time and Direction of Arrival Estimation of Stray Signals in a RCS/Antenna Range," Proc. of Antenna Measurement Techniques Annual Meeting and Symposium, Seattle, WA, October 1996.
33. W. D. Burnside, I. J. Gupta, "A Method to Reduce Stray Signal Errors in Antenna Pattern Measurements," *IEEE Transactions on Antennas and Propagation*, vol. 42, no. 3 (March 1994).

P · A · R · T · 4

Topics Associated with Antennas

Chapter 51

Transmission Lines and Waveguides

Roderic V. Lowman

Techni-Quest, Inc.,

Rainee N. Simons*

Consultant

CONTENTS

51.1	GENERAL	51-4
51.2	OPEN-WIRE TRANSMISSION LINES	51-7
51.3	WIRES IN VARIOUS ENCLOSURES	51-8
51.4	PRINTED TRANSMISSION LINES	51-9
51.5	COAXIAL LINES: SOLID CONDUCTOR	51-20
51.6	FLEXIBLE COAXIAL LINES	51-24
51.7	HOLLOW-TUBE WAVEGUIDES	51-24
51.8	MISCELLANEOUS TYPES OF WAVEGUIDES	51-43
51.9	LIST OF SYMBOLS	51-51

*The author is with the NASA Glenn Research Center, Cleveland, OH.

51.1 GENERAL

In the applications of antennas, it is necessary to use some form of transmission line to connect the antenna to a transmitter or receiver. It is essential to know the propagation characteristics of the more common forms of line.

The transmission lines in Sections 51.2 to 51.6 are almost always used in the transverse electromagnetic (TEM) mode of propagation and therefore possess a unique characteristic impedance. In air, the guide wavelength λ_g is equal to the free-space wavelength λ . Propagation in this mode exists at any frequency, although above a certain frequency higher modes may also exist. It is assumed that the line loss is small and may be neglected in calculating the characteristic impedance.¹

The waveguides considered in Sections 51.7 and 51.8 use modes of propagation having a longitudinal component and do not possess a unique characteristic impedance.² Propagation may only take place above a certain unique cutoff frequency defined by the mechanical dimensions. The guide wavelength λ_g is related to the free-space wavelength λ by $1/\lambda_g^2 = 1/\lambda^2 - 1/\lambda_c^2$ where λ_c is the cutoff wavelength.

In all cases, it will be assumed that the skin depth is small compared with the dimensions of the conductors and that these conductors are nonmagnetic. The medium surrounding the conductors is often air, but for other dielectrics a loss will be introduced, and this is independent of the line dimensions. In the case of TEM modes this loss is

$$\alpha_D = \frac{\pi}{\lambda_D} \tan \delta \quad \text{Np/m}$$

where $\lambda_D (= \lambda / \sqrt{\epsilon})$ is the intrinsic wavelength in the medium and $\tan \delta$ is the loss tangent of the dielectric. ϵ_r is the relative dielectric constant. For waveguide modes, the wavelength in the guide must be considered, and the dielectric loss becomes

$$\alpha_D = \frac{\pi \lambda_g}{(\lambda_D)^2} \tan \delta \quad \text{Np/m}$$

In the circuits that have discontinuities, such as bends, the open types of lines will have additional loss because of radiation at these points, and therefore shielded types should generally be used for complex circuits.

The power-handling capacity of transmission lines in the following sections is given in terms of the maximum allowable field intensity E_a in the dielectric, whereas for waveguides the power is given directly for air dielectric. For such calculations, $E_a = 3 \times 10^6$ V/m is the theoretical maximum for air dielectric, at normal temperature and atmospheric pressure, but for proper derating a value of 2×10^6 V/m is more practical. Other dielectrics having a higher dielectric strength may be used to increase the power limit, but in the case of solid dielectrics, the increased field strength in air pockets (which are very likely to exist) imposes a serious limitation. A common method of increasing the power limit is to use pressurized air, the maximum power being proportional to the square of the absolute pressure.

The conditions discussed above apply to a matched line. The maximum power is inversely proportional to the standing-wave ratio.

Important transmission-line equations are shown in Table 51-1 for both TEM and non-TEM lines and waveguides.

TABLE 51-1 Transmission-Line Equations

Quantity	General Line		Ideal Line		Approximate Results for Low-loss Lines	
	TEM Line	Other	TEM	Other	TEM Line	Other
Propagation constant $\gamma = \alpha + j\beta$	$\sqrt{(R + j\omega L)(G + j\omega C)}$		$j\omega\sqrt{LC}$	$j\frac{2\pi}{\lambda_g}$	(See α and β below)	(See α and β)
Phase constant β	$\text{Im}(\gamma)$	$\text{Im}(\gamma)$	$\omega\sqrt{LC} = \frac{\omega}{v} = \frac{2\pi}{\lambda}$	$\frac{2\pi}{\lambda_g}$	$\omega\sqrt{LC} \left[1 - \frac{RG}{4\omega^2 LC} + \frac{G^2}{8\omega^2 C^2} + \frac{R^2}{8\omega^2 L^2} \right]$	$\frac{2\pi}{\lambda_g}$
Attenuation constant α	$\text{Re}(\gamma)$	$\text{Re}(\gamma)$	0	0	$\frac{R}{2Z_0} + \frac{GZ_0}{2} \approx \frac{\pi}{\lambda} \tan \delta^*$	$\frac{\pi}{\lambda} \lambda_g \tan \delta^*$
Characteristic impedance Z_0	$\sqrt{\frac{R + j\omega L}{G + j\omega C}}$	Geometry-dependent	$\frac{L}{C}$	Geometry-dependent	$\sqrt{\frac{L}{C}} \left[1 + j \left(\frac{G}{2\omega C} - \frac{R}{2\omega L} \right) \right]$	Geometry-dependent
Input impedance Z_i	$Z_0 \left[\frac{Z_L \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_L \sinh \gamma l} \right]$		$Z_0 \left[\frac{Z_L \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_L \sin \beta l} \right]$		$Z_0 \left[\frac{\alpha l \cos \beta l + j \sin \beta l}{\cos \beta l + j\alpha l \sin \beta l} \right]$	
Impedance of shorted line Impedance of open line	$Z_0 \tanh \gamma l$ $Z_0 \coth \gamma l$		$jZ_0 \tan \beta l$ $-jZ_0 \cot \beta l$		$Z_0 \left[\frac{\cos \beta l + j\alpha l \sin \beta l}{\alpha l \cos \beta l + j \sin \beta l} \right]$	
Impedance of quarter-wave line	$Z_0 \left[\frac{Z_L \sinh \alpha l + Z_0 \cosh \alpha l}{Z_0 \sinh \alpha l + Z_L \cosh \alpha l} \right]$		$\frac{Z_0^2}{Z_L}$		$Z_0 \left[\frac{Z_0 + Z_L \alpha l}{Z_L + Z_0 \alpha l} \right]$	
Impedance of half-wave line	$Z_0 \left[\frac{Z_L \cosh \alpha l + Z_0 \sinh \alpha l}{Z_0 \cosh \alpha l + Z_L \sinh \alpha l} \right]$		Z_L		$Z_0 \left[\frac{Z_L + Z_0 \alpha l}{Z_0 + Z_L \alpha l} \right]$	

TABLE 51-1 Transmission-Line Equations (continued)

Quantity	General Line		Ideal Line		Approximate Results for Low-loss Lines	
	TEM Line	Other	TEM	Other	TEM Line	Other
Voltage along line $V(z)$	$V_i \cosh \gamma z - I_i Z_0 \sinh \gamma z$		$V_i \cos \beta z - j I_i Z_0 \sin \beta z$			
Current along line $I(z)$	$I_i \cosh \gamma z - \frac{V_i}{Z_0} \sinh \gamma z$		$I_i \cos \beta z - j \frac{V_i}{Z_0} \sin \beta z$			
Reflection coefficient ρ	$\frac{Z_L - Z_0}{Z_L + Z_0}$		$\frac{Z_L - Z_0}{Z_L + Z_0}$			
Standing-wave ratio	$\frac{1 + \rho }{1 - \rho }$		$\frac{1 + \rho }{1 - \rho }$			

Symbols:

R, L, G, C = distributed resistance, inductance, conductance, and capacitance per unit length

l = length of line.

Subscript i denotes input-end quantities

Subscript L denotes load-end quantities.

z = distance along line from input end

λ = wavelength measured in dielectric

v = phase velocity of line equals velocity of light in dielectric of line for an ideal line

λ_g = guide wavelength

$\tan \delta$ = loss tangent of dielectric

*Case of low conductor loss

51.2 OPEN-WIRE TRANSMISSION LINES

Open-wire transmission lines are an arrangement of wires whose diameters generally are small compared with the spacings involved. This arrangement is sometimes used in conjunction with a ground plane to which the wires are parallel. Such transmission lines have the advantage of simplicity and economy. The spacing between the wires and between the wires and the ground plane is very much less than a wavelength.

The most common of the open-wire lines is the two-wire line. For two wires of diameter d , spaced at a center-to-center distance D , the characteristic impedance is

$$Z_0 = \frac{120}{\sqrt{\epsilon}} \cosh^{-1} \frac{D}{d} \quad \Omega$$

This relation is plotted in Figure 51-1. In the case of unequal wire diameters d_1 and d_2 , d is replaced by $\sqrt{d_1 d_2}$.

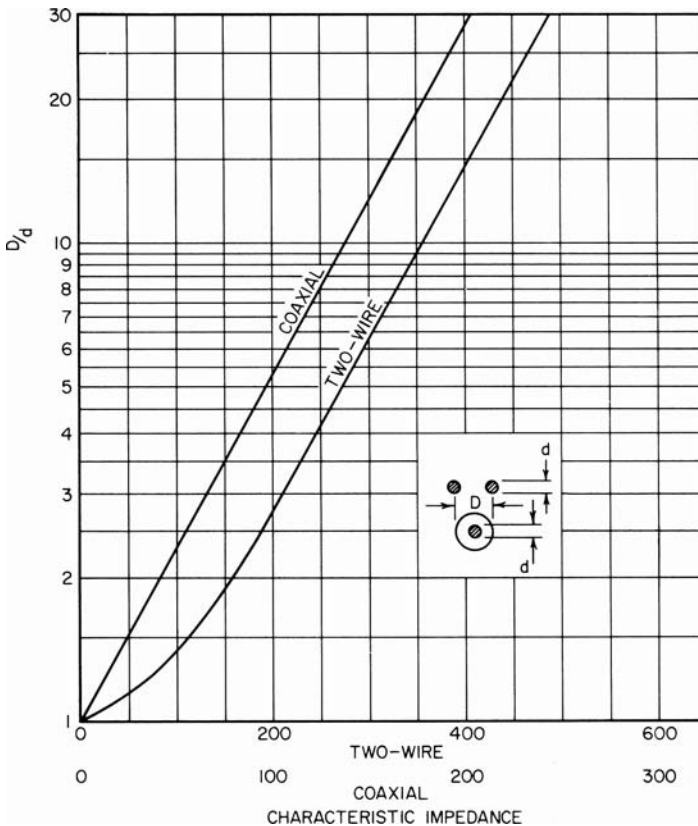


FIGURE 51-1 Characteristic impedance of common lines

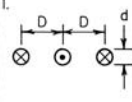
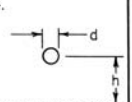
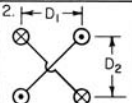
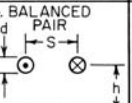
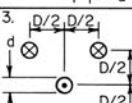
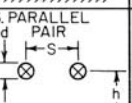
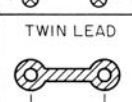
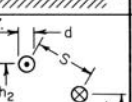
WIRES IN SPACE	CHARACTERISTIC IMPEDANCE	WIRES NEAR GROUND	CHARACTERISTIC IMPEDANCE
1. 	$Z_0 = \frac{207}{\sqrt{\epsilon}} \log_{10} 1.59 \frac{D}{d}$	4. 	$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} \frac{4h}{d}$
2. 	$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} \frac{2D_2}{d\sqrt{1+(D_2/D_1)^2}}$	5. BALANCED PAIR 	$Z_0 = \frac{276}{\sqrt{\epsilon}} \log_{10} \frac{4h}{d\sqrt{1+(2h/S)^2}}$
3. 	$Z_0 = \frac{173}{\sqrt{\epsilon}} \log_{10} 1.14 \frac{D}{d}$	6. PARALLEL PAIR 	$Z_0 = \frac{69}{\sqrt{\epsilon}} \log_{10} \frac{4h}{d\sqrt{1+(2h/S)^2}}$
TWIN LEAD 	Z_0 DEPENDS ON WIRE DIAMETER, SPACING, AND THE SHAPE OF THE DIELECTRIC, USUALLY POLYETHYLENE. IT IS MOSTLY MADE IN 300 Ω .	7. 	$Z_0 = \frac{276}{\sqrt{\epsilon}} \log_{10} \frac{2S}{d\sqrt{1+S^2/4h_1h_2}}$

FIGURE 51-2 Open-wire lines (Note that d is small compared with other dimensions.)

The conductor loss is

$$\alpha_c = 2.29 \times 10^{-6} \sqrt{\frac{f\epsilon}{\sigma}} \frac{1}{d \log_{10}(2D/d)} \quad \text{Np/m}$$

The power-handling capacity is

$$P = \frac{E_a^2 d^2 \sqrt{\epsilon}}{240} \cosh^{-1} \frac{D}{d} \quad \text{W}$$

where E_a is the maximum allowable field intensity in the dielectric (Section 51.1).

Other configurations of open-wire lines are shown in Figure 51-2, together with formulas for the characteristic impedance. Lines near a ground plane are also included. Other types of open-wire lines can be found in the literature.

51.3 WIRES IN VARIOUS ENCLOSURES

It is often advantageous to tailor a transmission line to fit a specific application when one of the more common forms is not convenient. Furthermore, with many antenna configurations a nonstandard form of line has definite advantages.

Various forms of transmission line comprising wires in different-shaped enclosures are shown in Figure 51-3, together with formulas for characteristic impedance. These lines are shielded, with the exception of the corner enclosure, provided that those that are physically open have an opening less than one-half wavelength across, extend beyond the wire at least one-half wavelength, and have opposing surfaces that are maintained at the same potential.

SINGLE WIRE	CHARACTERISTIC IMPEDANCE	BALANCED WIRE PAIR	CHARACTERISTIC IMPEDANCE
1.	$Z_0 = \frac{138}{\sqrt{\epsilon}} \left[\log_{10} \frac{4D}{\pi d} - \frac{0.0367 \left(\frac{d}{D} \right)^2}{1 - 0.355 \left(\frac{d}{D} \right)^4} \right]$	5.	$Z_0 = \frac{276}{\sqrt{\epsilon}} \log_{10} \frac{4D}{\pi d} \tanh \frac{\pi S}{2D}$
2.	$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} \frac{\sqrt{2} D}{d}$	6.	$Z_0 = \frac{276}{\sqrt{\epsilon}} \log_{10} \frac{2D}{\pi d}$
3.	$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} \frac{4D}{\pi d} \tanh \frac{\pi h}{D}$ FOR EXACT FORMULA SEE REF. 3	7.	$Z_0 = \frac{276}{\sqrt{\epsilon}} \log_{10} \frac{2D}{\pi d \sqrt{A}}$ $A = \csc^2 \left(\frac{\pi S}{d} \right) + \csc^2 \left(\frac{2\pi h}{d} \right)$
4.	$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} 1.08 \frac{D}{d}$	8.	$Z_0 = \frac{276}{\sqrt{\epsilon}} \left(\log_{10} \frac{4D}{\pi d} \tanh \frac{\pi S}{2D} - \sum_{n=1}^{\infty} \log_{10} \frac{1+u_n^2}{1-v_n^2} \right)$ $u_n = \frac{\sinh \frac{\pi S}{2D}}{\cosh \frac{n\pi W}{2D}}, \quad v_n = \frac{\sinh \frac{\pi S}{2D}}{\sinh \frac{n\pi W}{2D}}$

FIGURE 51-3 Wires in various enclosures³ (Note that d is small compared with other dimensions except in case 1.)

51.4 PRINTED TRANSMISSION LINES

In printed transmission lines, the conductors are flat strips that most frequently are photoetched from a dielectric sheet which is copper-clad on one or both sides. Although not often used as uniform transmission lines, they are very useful for antenna feeds, networks, printed arrays, and slots. They are also useful for a number of microwave integrated circuits such as amplifiers, filters, and high-speed digital logic circuits.

A number of different types of these transmission lines have been suggested, but a good variety of useful characteristics are supplied by six basic types that complement each other well: microstrip, suspended microstrip, inverted microstrip, slotline, coplanar waveguide, and coplanar strip.⁴ Cross sections of these types are shown in Figure 51-4, and closed form equations for computing Z_0 and λ_g or effective dielectric constant are given in the subsections that follow.

Microstrip is made of a dielectric sheet with a narrow conducting pattern of copper on one side and a copper sheet serving as a ground plane on the other. Transmission is mostly of the TEM mode with a good part of the electric field between the broad face of the conductor and the large ground plane. Microstrip is similar to a wire above ground that has been flattened. The field pattern is more complex, as only a part of it is in the dielectric sheet and the rest is in the air above the dielectric sheet. The other three are also mostly TEM, but their electric field is between the edges of the conductor pattern.

Coplanar stripline is similar to a two-wire line with flattened wires. It too is complex, with the fields partly in the dielectric and partly in the air above and below the sheet. The impedance is high because the capacity per unit length is low owing to the edge-to-edge position of the conductors. Coplanar waveguide is really two coplanar striplines in parallel.

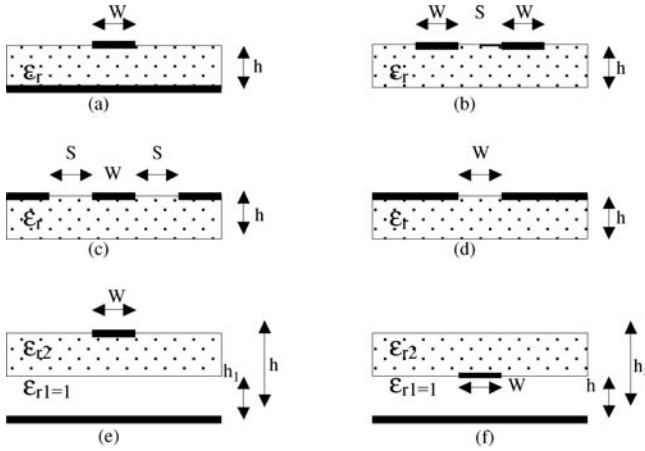


FIGURE 51-4 Printed transmission lines: (a) microstrip, (b) coplanar strips; (c) coplanar waveguide; (d) slotline; (e) suspended microstrip; (f) inverted microstrip

As might be expected, the impedance is lower for the same dimensions and dielectric. Slotline is similar to coplanar stripline, with very wide conductor patterns.

Microstrip Line

For a microstrip line of width W on a dielectric substrate of thickness h and relative dielectric constant ϵ_r , the characteristic impedance Z_0 in ohms is given by the expression⁵:

$$Z_0 = \frac{\eta_0}{2.0\sqrt{2.0\pi\sqrt{\epsilon_r+1.0}}} \ln \left\{ 1.0 + \frac{4.0h}{w'} \left[\frac{14.0+8.0/\epsilon_r}{11.0} - \frac{4.0h}{w'} \right] + \sqrt{\left(\frac{14.0+8.0/\epsilon_r}{11.0} \right)^2 \left(\frac{4.0h}{w'} \right)^2 + \frac{1.0+1.0/\epsilon_r}{2.0} \pi^2} \right\}$$

where w' is an equivalent strip width that takes into account the thickness t of the strip and is given by

$$w' = W + \Delta w'$$

$$\Delta w' = \Delta w \left(\frac{1.0+1.0/\epsilon_r}{2.0} \right)$$

$$\frac{\Delta w}{t} = \frac{1.0}{\pi} \ln \left[\frac{4e}{\sqrt{\left(\frac{t}{h} \right)^2 + \left(\frac{1}{\frac{W}{t} + 1.1} \right)^2}} \right]$$

For $W/h \leq 1.0$, the effective dielectric constant ϵ_{eff} is given by the expression:

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[\left(1 + \frac{12h}{W} \right)^{-0.5} + 0.04 \left(1.0 - \frac{W}{h} \right)^2 \right]$$

For $W/h \geq 1$:

$$\epsilon_{eff} = \frac{\epsilon_r + 1.0}{2.0} + \frac{\epsilon_r - 1.0}{2.0} \left(1 + \frac{12.0h}{W} \right)^{-0.5}$$

The previous equations are accurate to within 1% for

$$\epsilon_r \leq 16 \quad \text{and} \quad 0.05 \leq \frac{W}{h} \leq 20.0$$

and within 2% for

$$\epsilon_r > 16 \quad \text{and} \quad \frac{W}{h} < 0.05$$

The error in Z_0 is less than 2% for any ϵ_r and W .

Coplanar Stripline

For a symmetric coplanar stripline with strip width W and separation S on a dielectric substrate of thickness h and relative dielectric constant ϵ_r , the characteristic impedance Z_0 in ohms is given by the expression:⁵

$$Z_0 = \frac{\eta_0}{\sqrt{\epsilon_{eff}}} \frac{K(k)}{K(k')}$$

where the effective dielectric constant ϵ_{eff} is given by

$$\epsilon_{eff} = 1 + \frac{\epsilon_r - 1}{2} \frac{K(k')}{K(k)} \frac{K(k_1)}{K(k'_1)}$$

$$k = \frac{S}{S + 2W}$$

$$k' = \sqrt{1.0 - k^2}$$

$$k_1 = \frac{\sinh\left(\frac{\pi S}{4h}\right)}{\sinh\left(\frac{\pi(S + 2W)}{4h}\right)}$$

$$k'_1 = \sqrt{1.0 - k_1^2}$$

The ratio of the complete elliptic integrals of the first kind can be approximated by the following expressions, for $1.0 \leq K(k)/K(k') \leq \infty$ and $1.0/\sqrt{2} \leq k \leq 1.0$:

$$\frac{K(k)}{K(k')} = \frac{1.0}{2.0\pi} \ln \left[2.0 \frac{\sqrt{1.0+k} + \sqrt[4]{4.0k}}{\sqrt{1.0+k} - \sqrt[4]{4.0k}} \right]$$

For $0 \leq K(k)/K(k') \leq 1.0$ and $0 \leq k \leq 1.0/\sqrt{2}$:

$$\frac{K(k)}{K(k')} = \frac{2.0\pi}{\ln \left[2.0 \frac{\sqrt{1.0+k'} + \sqrt[4]{4.0k'}}{\sqrt{1.0+k'} - \sqrt[4]{4.0k'}} \right]}$$

In deriving the previous expressions the thickness t is assumed to be zero.

Coplanar Waveguide

For a symmetric coplanar waveguide with center strip conductor of width W and separated by slots of width S from adjacent ground planes, on a dielectric substrate of thickness h and relative dielectric constant ϵ_r , the characteristic impedance Z_0 in ohms is given by the expression:⁵

$$Z_0 = \frac{30.0\pi}{\sqrt{\epsilon_{eff,t}}} \frac{K(k_t')}{K(k_t)}$$

where $\epsilon_{eff,t}$ is the conductor thickness t dependent effective dielectric constant and is given by

$$\epsilon_{eff,t} = \epsilon_{eff} - \frac{\epsilon_{eff} - 1.0}{\frac{(b-a)}{1.4t} \frac{K(k)}{K(k')} + 1.0}$$

$$\epsilon_{eff} = 1.0 + \frac{\epsilon_r - 1.0}{2.0} \frac{K(k')}{K(k)} \frac{K(k_1)}{K(k_1')}$$

$$k = \frac{a}{b}, \quad k' = \sqrt{1.0 - k^2}, \quad a = W \quad \text{and} \quad b = W + 2.0S$$

$$k_1 = \frac{\sinh\left(\frac{\pi a_t}{4.0h}\right)}{\sinh\left(\frac{\pi b_t}{4.0h}\right)}, \quad k_1' = \sqrt{1.0 - k_1^2}$$

$$k_t = \frac{a_t}{b_t}, \quad k_t' = \sqrt{1.0 - k_t^2}$$

$$a_t = a + \frac{1.25t}{\pi} \left[1.0 + \ln \left(\frac{4.0\pi a}{t} \right) \right]$$

$$b_t = b - \frac{1.25t}{\pi} \left[1.0 + \ln \left(\frac{4.0\pi a}{t} \right) \right]$$

The ratio $K(k)/K(k')$ can be determined using the expressions given in the previous section on coplanar striplines. Coplanar waveguide also has application in the design of

components such as directional couplers, filters, and baluns. More information on these components and systems can be found in the reference section.⁶

Slotline

In antenna applications, slotlines are fabricated either on low relative dielectric constant plastic substrates with $\epsilon_r \approx 2.2$ or on high relative dielectric constant ceramic substrates with $\epsilon_r \approx 9.8$. For a slotline with slot width W on a dielectric substrate of thickness h and relative dielectric constant ϵ_r , the characteristic impedance Z_0 in ohms at a free space wavelength λ_0 is given by the expression:⁷

For $0.0015 \leq W/\lambda_0 \leq 0.075$ and $2.2 \leq \epsilon_r \leq 3.8$

$$\begin{aligned} Z_0 = & 60.0 + 3.69 \sin \left[\frac{(\epsilon_r - 2.22)\pi}{2.36} \right] + 133.5 \ln(10.0 \epsilon_r) \sqrt{W/\lambda_0} \\ & + 2.81[1.0 - 0.011 \epsilon_r (4.48 + \ln \epsilon_r)](W/h) \ln(100.0 h/\lambda_0) \\ & + 131.1(1.028 - \ln \epsilon_r) \sqrt{h/\lambda_0} \\ & + 12.48(1.0 + 0.18 \ln \epsilon_r) \frac{W/h}{\sqrt{\epsilon_r - 2.06 + 0.85(W/h)^2}} \end{aligned}$$

The corresponding effective dielectric constant ϵ_{eff} at a free space wavelength λ_0 is given by

$$\epsilon_{eff} = \left\{ 1.045 - 0.365 \ln \epsilon_r + \frac{6.4(W/h) \epsilon_r^{0.945}}{238.64 + 100.0 W/h} - \left[0.148 - \frac{8.81(\epsilon_r + 0.95)}{100.0 \epsilon_r} \right] \ln \frac{h}{\lambda_0} \right\}^{-2}$$

The wavelength in the slotline or the guide wavelength λ_g can be calculated from ϵ_{eff} using the expression

$$\epsilon_{eff} = \left(\frac{\lambda_g}{\lambda_0} \right)^{-2}$$

Using the previous set of expressions, the computed Z_0 and λ_g/λ_0 for $0.006 \leq h/\lambda_0 \leq 0.060$ are accurate to better than 2.7% and 2.2%, respectively.

For $0.075 \leq W/\lambda_0 \leq 1.0$ and $2.2 \leq \epsilon_r \leq 3.8$

$$\begin{aligned} Z_0 = & 133.0 + 10.34(\epsilon_r - 1.0)^2 + 2.87[2.96 + (\epsilon_r - 1.582)^2] \\ & \left\{ (W/h + 2.32 \epsilon_r - 0.56) [(32.5 - 6.67 \epsilon_r)(100.0 h/\lambda_0)^2 - 1.0] \right\}^{0.5} \\ & - (684.45 h/\lambda_0)(\epsilon_r + 1.35)^2 + 13.23 [(\epsilon_r - 1.722)W/\lambda_0]^2 \end{aligned}$$

The corresponding effective dielectric constant ϵ_{eff} at a free space wavelength λ_0 is given by

$$\epsilon_{eff} = \left\{ 1.194 - 0.24 \ln \epsilon_r - \frac{0.621 \epsilon_r^{0.835} (W/\lambda_0)^{0.48}}{1.344 + W/h} - 0.0617 \left[1.91 - \frac{\epsilon_r + 2.0}{\epsilon_r} \right] \ln(h/\lambda_0) \right\}^{-2}$$

The computed Z_0 and λ_g/λ_0 for $0.006 \leq h/\lambda_0 \leq 0.060$ are accurate to better than 5.4% and 2.6%, respectively.

For $0.0015 \leq W/\lambda_0 \leq 0.075$ and $3.8 \leq \epsilon_r \leq 9.8$

$$\begin{aligned}
 Z_0 &= 73.6 - 2.15 \epsilon_r + (638.9 - 31.37 \epsilon_r)(W/\lambda_0)^{0.6} \\
 &\quad + \left(36.23\sqrt{\epsilon_r^2 + 41.0} - 225.0\right) \frac{W/h}{W/h + 0.876 \epsilon_r - 2.0} \\
 &\quad + 0.51(\epsilon_r + 2.12)(W/h) \ln\left(\frac{100.0h}{\lambda_0}\right) - 0.753 \epsilon_r \frac{h/\lambda_0}{\sqrt{W/\lambda_0}} \\
 \epsilon_{eff} &= \left\{ 0.9217 - 0.277 \ln \epsilon_r + 0.0322 \left(\frac{W}{h} \right) \sqrt{\frac{\epsilon_r}{\frac{W}{h} + 0.435}} - 0.01 \ln \left(\frac{h}{\lambda_0} \right) \left[4.6 - \frac{3.65}{\epsilon_r^2 \sqrt{\frac{W}{\lambda_0} \left(9.06 - 100 \frac{W}{\lambda_0} \right)}} \right] \right\}^{-2}
 \end{aligned}$$

The computed Z_0 and λ_g/λ_0 for $0.006 \leq h/\lambda_0 \leq 0.060$ are accurate to better than 5.4% and 3.0%, respectively.

For $0.075 \leq W/\lambda_0 \leq 1.0$ and $3.8 \leq \epsilon_r \leq 9.8$

$$\begin{aligned}
 Z_0 &= 120.75 - 3.74 \epsilon_r + 50.0 \left[\tan^{-1}(2.0 \epsilon_r) - 0.8 \right] \\
 &\quad \times \left(\frac{W}{h} \right)^{\left\{ 1.11 + \left[0.132(\epsilon_r - 27.7) / \left(100.0 \frac{h}{\lambda_0} + 5.0 \right) \right] \right\}} \\
 &\quad \times \ln \left[100 \frac{h}{\lambda_0} + \sqrt{\left(100 \frac{h}{\lambda_0} \right)^2 + 1.0} \right] \\
 &\quad + 14.21(1.0 - 0.458 \epsilon_r) \left(100.0 \frac{h}{\lambda_0} + 5.1 \ln \epsilon_r - 13.1 \right) \left(\frac{W}{\lambda_0} + 0.33 \right)^2 \\
 \epsilon_{eff} &= \left\{ 1.05 - 0.04 \epsilon_r + 1.411 \times 10^{-2} (\epsilon_r - 1.421) \right. \\
 &\quad \times \ln \left[\frac{W}{h} - 2.012 (1.0 - 0.146 \epsilon_r) \right] \\
 &\quad + 0.111 (1.0 - 0.366 \epsilon_r) \sqrt{\frac{W}{\lambda_0}} \\
 &\quad \left. + 0.139 [1.0 + 0.52 \epsilon_r \ln(14.7 - \epsilon_r)] \left(\frac{h}{\lambda_0} \right) \ln \left(\frac{h}{\lambda_0} \right) \right\}^{-2}
 \end{aligned}$$

The computed Z_0 and λ_g/λ_0 for $0.006 \leq h/\lambda_0 \leq 0.060$ are accurate to better than 5.8% and 3.2%, respectively.

A low loss transition between a slotline and a fin line is very simple and easy to design and therefore used extensively to excite a tapered slot antenna (TSA) at millimeter wave frequencies.⁸

The ranges of Z_0 available and the losses with the different types of strip transmission lines are compared in Table 51-2. These show that the microstrip and the coplanar waveguide

TABLE 51-2 Comparison of Z_0 Limits and Loss for the Various Lines ($\epsilon_r = 10.0$, $h = 25$ mil, or 0.635 mm, and frequency = 10 GHz)

Transmission line	Z_0 range		Loss, dB/cm	
	Minimum	Maximum	50	100
Microstrip	20 (m)	110 (d)	0.04	0.14
Coplanar waveguide	25 (m, d)	155 (m, d)	0.08*	0.28*
Slotline	55 (d)	300 (m)	0.15†	
Coplanar strips	45 (m, d)	280 (m, d)	0.83*	0.13*

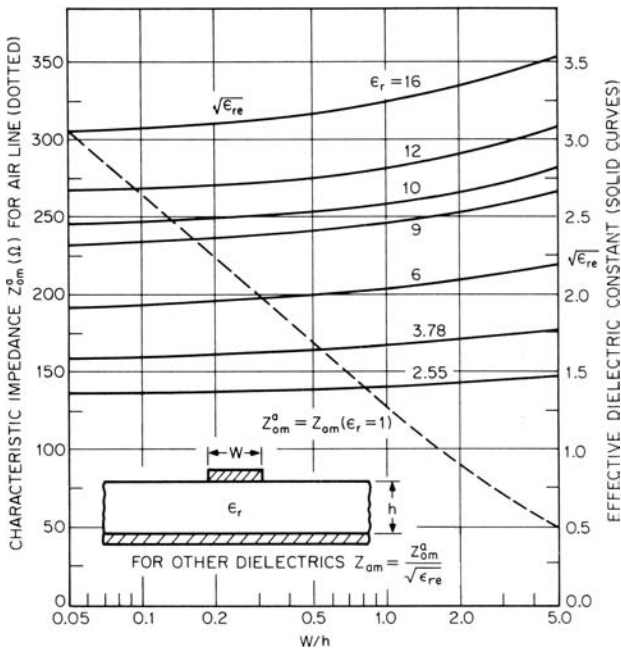
NOTE: (m) = Z_0 limited by mode; (d) = Z_0 limited by small dimensions

* $h/W = 2$

† $\epsilon_r = 16$, $Z_0 = 75 \Omega$

are most useful for low-impedance circuits and that the slotline and coplanar strips are more appropriate for high-impedance circuits.⁹

The characteristics of microstrip are shown in Figure 51-5. The dotted curve is the plot of Z_{0m}^a for airline, while the solid curves are for finding the square root of the effective dielectric constant. The Z_{0m} for other dielectrics is $Z_{0m}^a / \sqrt{\epsilon_{re}}$. Figure 51-6 shows the losses in microstrip on different substrates. Line wavelengths and characteristic impedance of coplanar strips and of coplanar waveguide are shown in Figures 51-7 and 51-8. These same characteristics for slotline are shown in Figure 51-9.

**FIGURE 51-5** Characteristic impedance and effective dielectric constant of microstrip lines

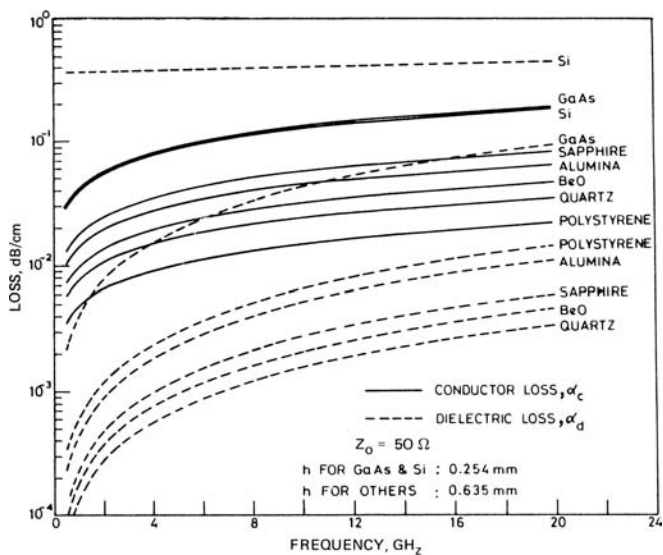


FIGURE 51-6 The conductor and dielectric losses when microstrip is constructed on different substrates

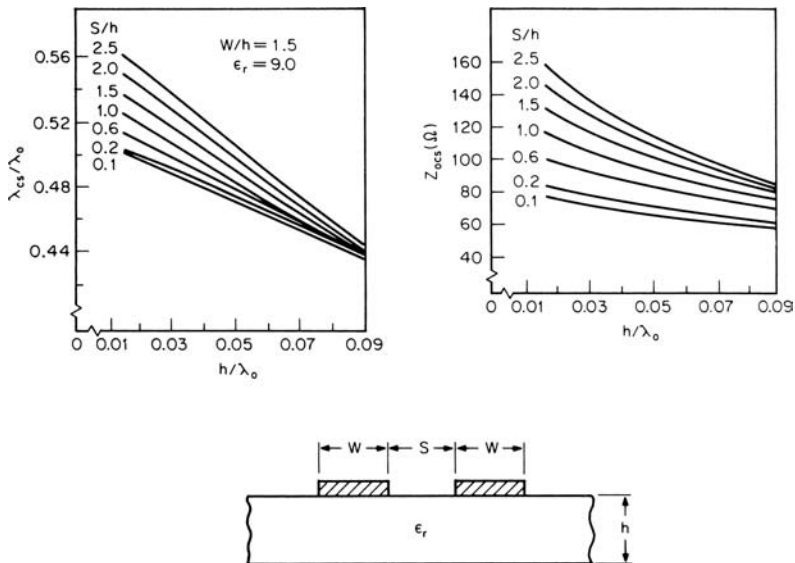


FIGURE 51-7 Line wavelength and characteristic impedance of coplanar strips

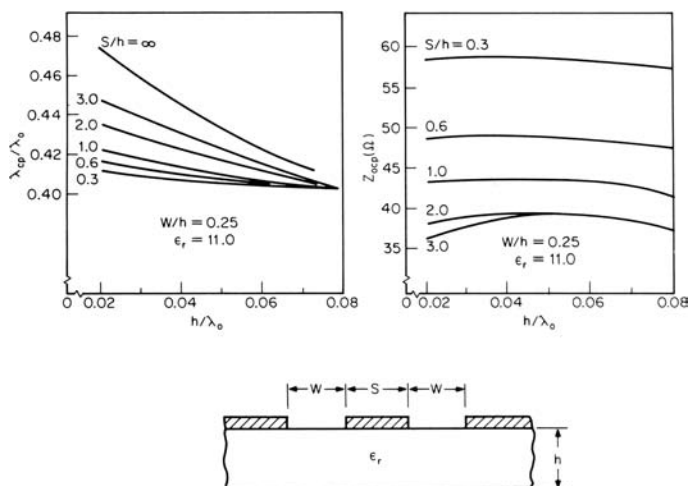


FIGURE 51-8 Line wavelength and characteristic impedance of coplanar waveguide

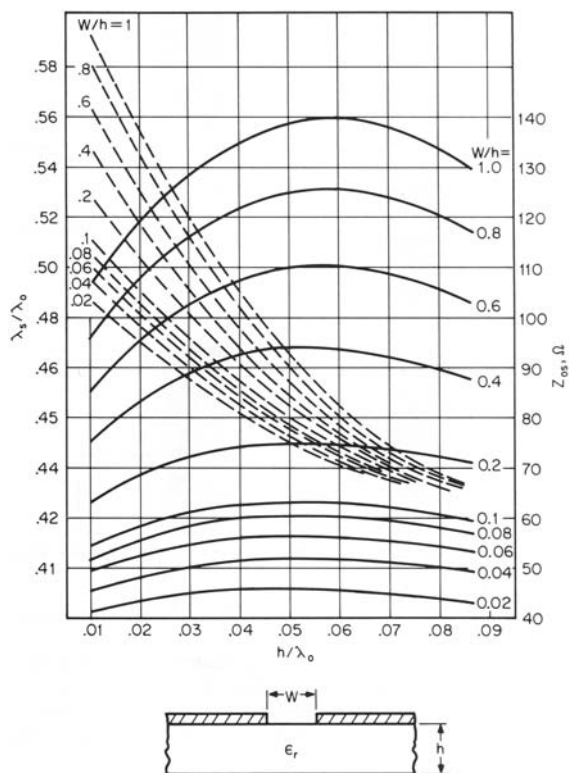


FIGURE 51-9 Line wavelength and characteristic impedance of slotline

Suspended Microstrip

The suspended and inverted microstrip lines have lower dispersion than conventional microstrip lines. Hence, simple quasi-TEM mode analysis yields quite satisfactory results for components such as couplers and filters at higher microwave frequencies. In addition, for the same characteristic impedance, substrate thickness, and a comparable air gap, the strip width when compared to conventional microstrip line is larger—which lowers the attenuation due to conductor losses by a factor of 2 to 3 and also eases fabrication tolerances.

The effective dielectric constant ϵ_{eff} for a suspended microstrip line with strip width W , air gap height h_1 , substrate thickness $h - h_1$, and relative dielectric constant ϵ_{r2} is given by¹⁰

$$\epsilon_{eff} = 1 - q_1 - q_2 + \epsilon_{r1}\epsilon_{r2}(q_1 + q_2)^2/(\epsilon_{r1}q_2 + \epsilon_{r2}q_1)$$

For wide lines with $W/h > 1$, the filling factors are given by

$$q_1 = f_1 \{1 - 0.6523 / u_1 \ln[1.801 u_1 \sin(\phi)/x + f_2 \cos(\phi)]\}$$

$$q_2 = 1 - q_1 - q_w$$

In the previous expression q_w is the Wheeler's filling factor for wide lines and is given by

$$q_w = 0.6523/u_1 \ln(1.801 u_1)$$

In addition, u_1 is the normalized effective line width given by

$$u_1 = w_{eff}/h = u + 2/\pi \ln[17.08(u/2 + 0.92)]$$

$$u = W/h$$

$$x = h_1/h$$

$$\phi = \frac{\pi}{2} x$$

$$f_1 = x^{(0.8605 + 0.3491/E)}$$

$$E = \epsilon_{r2}/\epsilon_{r1}$$

$$f_2 = (1.929 + 5.908/E)u_1^\alpha$$

$$\alpha = (1.208 + 0.1077E) \times (u/4)^{(0.3048(0.5-x))}$$

For narrow lines with $W/h \leq 1$, the filling factors are

$$q_1 = q_n \left\{ 1 - \left[2/\pi \cos^{-1}(x f_1) \right]^{f_2} \right\}$$

$$q_2 = q_n - q_1$$

where q_n is the Wheeler's filling factor for narrow lines given by

$q_n = 1/2 + 0.26144 / \ln(8/u) \times (1 + 0.2855E_{12}^{-0.7517})$ and x , u , and E are as defined previously

$$E_{12} = x(\epsilon_{r1} - \epsilon_{r2}) + \epsilon_{r2}$$

$$f_1 = [u_1/2(1+x)/(1-x+u_1)]^{0.3915}$$

$$u_1 = (0.13955 + 0.8095/E)(0.0538 + 0.597u + 0.5624u^2)$$

$$f_2 = [1.677 + 0.7848/E - (1.4021 + 0.06055E) \times x + (0.7954 + 0.04382E) \times x^2] \times (u/0.4)^{[0.07(x-0.5)]}$$

Having determined the ϵ_{eff} , the characteristic impedance Z_0 in ohms is calculated as follows

$$Z_0 = \frac{Z}{(\epsilon_{eff})^{0.5}}$$

where Z is the characteristic impedance of an identical air-filled microstrip line and is given by

$$Z = \frac{\eta_0}{2\pi} \ln \left[\frac{f(u)}{u} + \sqrt{1 + \left(\frac{2}{u}\right)^2} \right]$$

where

$$f(u) = 6 + (2\pi - 6) \exp \left[- \left(\frac{30.666}{u} \right)^{0.7528} \right]$$

and $\eta_0 = 376.73 \Omega$ and $u = W/h$.

Inverted Microstrip

The effective dielectric constant ϵ_{eff} for an inverted microstrip line with strip width W , air gap height h , substrate thickness $h_2 - h$, and relative dielectric constant ϵ_{r2} is given by¹⁰

$$\epsilon_{eff} = \epsilon_{r1}q_1 + \epsilon_{r2}(1 - q_1)^2 / (\epsilon_{r2}q_3 + q_2)$$

For wide lines with $W/h > 1$, the filling factors are given by

$$q_1 = 1 - q_w f_1$$

$$q_2 = 1 - q_1 - q_3$$

$$q_3 = 0.6523(1 - f_2)u_1 \ln[u_1 \cos(\phi) / (0.5552 + f_2 \times f_3) + \sin(\phi)]$$

where u_1 and q_w are the normalized effective line width and Wheeler's filling factor for wide lines, respectively. The expressions for u_1 and q_w are identical to those given previously for the suspended microstrip line.

$$u = W/h$$

$$f_1 = 1 - 0.06078[1 - \cos(\pi x^{0.5})] \times [1 - 0.3206/u_1 - 2.3188/u_1^3] \times E^{0.4356}$$

$$x = 1 - h/h_2$$

$$E = \frac{\epsilon_{r2}}{\epsilon_{r1}}$$

$$f_2 = x^\alpha$$

$$\alpha = 0.4746 + 1.3778/E - (0.1376 + 0.00945E)x + 0.6153x^3$$

$$\phi = \frac{\pi}{2} f_2$$

$$f_3 = (3.9717 + 8.0922/E)u_1^{-(0.1321 + 1.3907/E)}$$

For $W/h \leq 1$, the filling factors are given by

$$q_1 = 0.5 + (q_n - 0.5) f_1$$

$$q_2 = 1 - q_1 - q_3$$

$$q_3 = (f_2 - f_3) / [\pi \ln(8/u)]$$

where q_n is the Wheeler's filling factor for narrow lines given by

$$q_n = 0.5 + 0.26144 / \ln(8/u) \times (1 + 0.2855 \epsilon_{r1}^{-0.7517})$$

and

$$f_1 = 1 + 0.1431[1 - \cos(\pi x)] \times [1 - 1.652 / \ln(12.532/u)] \times E^{0.641}$$

$$f_2 = \ln(A) \cos^{-1}[(x A^{0.5})^\alpha]$$

$$A = (2 - x) / [x + u/4(1 - x)]$$

$$\alpha = [0.14695 + 1.657/E + (0.6386 + 0.7881/E) \times x] \times (2x_1)^{[0.3075(u-0.4)]}$$

$$x_1 = 1 - x$$

$$f_3 = 1.0558 x_1^{1.928} (u/0.4042)^{[x(0.30035 + 0.2096/E)]}$$

Having determined the ϵ_{eff} , the Z_0 is determined in an analogous manner, using the expression given previously for suspended microstrip line.

51.5 COAXIAL LINES: SOLID CONDUCTOR

Coaxial transmission lines using a cylindrical center conductor within a cylindrical tubular outer conductor are widely employed for the propagation of microwave power. Although more costly than open-wire transmission lines, they completely enclose the electromagnetic fields, preventing radiation losses and providing shielding from nearby circuits.

Coaxial-Line Parameters

For a coaxial line with inner-conductor diameter $2a$ and outer conductor diameter $2b$, the characteristic impedance is

$$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} \frac{b}{a} \quad \Omega$$

which is plotted in Figure 51-1.

The cutoff wavelength of the first higher mode is²

$$\lambda_c = F\pi\sqrt{\epsilon} (a+b) \quad F \cong 1$$

The conductor loss for the dominant mode is

$$\alpha_c = 1.14 \times 10^{-6} \sqrt{\frac{f\epsilon}{\sigma}} \left(\frac{1}{a} + \frac{1}{b} \right) \frac{1}{\log_{10} \frac{b}{a}} \quad \text{Np/m}$$

The minimum conductor loss occurs at the limiting size for the first higher mode, in which case $\sqrt{\epsilon} Z_0 = 92.6 \Omega$ and

$$\alpha_c = 0.637 \frac{\epsilon}{(\lambda_c)^{3/2} \sqrt{\sigma}} \quad \text{Np/m}$$

For a fixed outer-conductor size, the minimum conductor loss occurs for $\sqrt{\epsilon} Z_0 = 77 \Omega$ and is

$$\alpha_c = 0.164 \frac{1}{b} \sqrt{\frac{\epsilon}{\sigma \lambda}} \quad \text{Np/m}$$

The power transmitted by the line is

$$P = \frac{E_a^2 a^2 \sqrt{\epsilon}}{52.2} \log_{10} \frac{b}{a} \quad \text{W}$$

where E_a is the electric field intensity at the center conductor (Section 51.1). The maximum power-handling capacity occurs for a 44.4Ω line operating at the limit of the first higher mode and is

$$P_{\max} = 6.53 \times 10^{-5} \frac{E_a^2 \lambda_c^2}{\sqrt{\epsilon}} \quad \text{W}$$

For a fixed outer-conductor size, the maximum power-handling capacity occurs for $\sqrt{\epsilon_r} Z_0 = 30 \Omega$ and is

$$P'_{\max} = 1.53 \times 10^{-3} E_a^2 b^2 \sqrt{\epsilon} \quad \text{W}$$

The effect of dimensional tolerances on the characteristic impedance may be found from

$$\frac{\Delta Z_0}{Z_0} = \frac{60}{Z_0 \sqrt{\epsilon}} \left(\frac{\Delta b}{b} - \frac{\Delta a}{a} \right)$$

while the effect of eccentricity is to change the characteristic impedance to

$$Z_0 = \frac{138}{\sqrt{\epsilon}} \log_{10} \left[\frac{b}{a} \left(1 - \frac{e^2}{b^2} \right) \right] \quad \Omega, \quad \frac{e}{b} < \log_{10} \frac{b}{a} < 0.7$$

where e is the off-center distance.

For a balanced coaxial line (sheathed two-wire line) having two center conductors spaced at a distance s within a single outer conductor, the characteristic impedance is given by

$$Z_0 = \frac{276}{\sqrt{\epsilon}} \log_{10} \left(\frac{s}{a} \frac{4b^2 - s^2}{4b^2 + s^2} \right) \quad \Omega$$

Coaxial Line with Helical Center Conductor

Coaxial lines with a helical inner conductor are sometimes useful to obtain a high characteristic impedance or slow propagation velocity. Design relations for this are given in the reference section.¹¹

Coaxial Line, Bead-Supported

In order to support the center conductor of an air-dielectric coaxial line, insulating beads are often used. These beads will introduce discontinuities and, if not properly designed, will all produce large reflection losses. The design procedure for these support beads depends upon the frequency range of application, a broad range requiring a more complex design. The more common procedures are summarized in Figure 51-10, and others can be found in the references.¹²

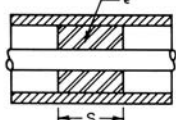

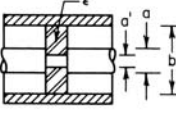
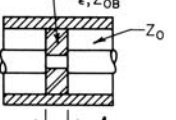
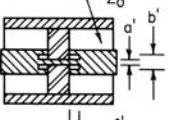
TYPE	CONFIGURATION	DESIGN FORMULAS	REMARKS
1. HALF WAVE		$S = \frac{\lambda}{2\sqrt{\epsilon}}$	FREQUENCY-SENSITIVE
2. SPACED PAIR		$S = \frac{1}{\beta} \arctan \left[\frac{2\sqrt{\epsilon}}{1+\epsilon} \cot \beta l \sqrt{\epsilon} \right]$ $\beta = \frac{2\pi}{\lambda}$	FREQUENCY-SENSITIVE
3. SIMPLE UNDERCUT		$a' = b \left(\frac{b}{a} \right)^{-\sqrt{\epsilon}}$	GOOD AT LOW FREQUENCY. NEGLECTS FRINGING CAPACITANCE AT UNDERCUT.
4. COMPENSATED UNDERCUT		$l = \frac{1}{\beta} \arctan \left[\frac{2\omega C Z_{0B}}{(\omega C Z_{0B})^2 + \left(\frac{Z_{0B}}{Z_0} \right)^2} \right]$ $\beta = \frac{2\pi\sqrt{\epsilon}}{\lambda}$	COMPENSATES FRINGING CAPACITANCE AT UNDERCUT. MATCH AT $f = \omega/2\pi$, LOW REFLECTION AT LOWER FREQUENCIES.
5. COMPOUND UNDERCUT (REF. 4)		$\sqrt{L/C} = Z_0$ $L = \frac{138 l'}{v} \log_{10} \frac{b'}{a'}$ FIND a' AS IN 3) ABOVE $v = 3 \times 10^8 \text{ m/s}$	VERY LOW REFLECTION IF DISCONTINUITIES ARE OF INFINITESIMAL EXTENT ($l' < \lambda/20$). GOOD WHERE MANY BEADS MUST BE USED.

FIGURE 51-10 Coaxial-line beads. Capacitive discontinuities at an abrupt change in diameter (as in items 4 and 5) can be found from this curve if the discontinuities are separated by at least the space between the inner and outer conductors in the intervening line. If the ratio of the inner and outer conductors is less than 5 before and after the discontinuity, the curves are accurate to better than 20%.

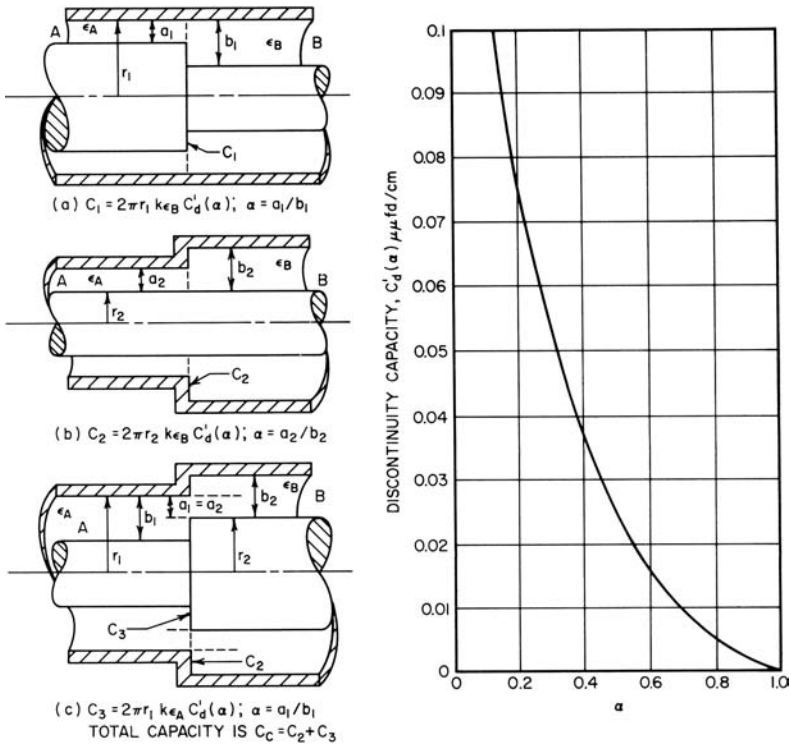


FIGURE 51-10 (Continued)

In the bead-supported line, the cutoff wavelength in the bead section should be kept below the operating wavelength. The power-handling capacity is about 0.033 times that of the theoretical maximum for the unsupported line.

Coaxial Line, Stub-Supported

In many applications, particularly when high power is to be transmitted, it is desirable to support the center conductor of an air-dielectric coaxial line by means of stubs. The power-handling capacity of a stub-supported line is about 0.15 times that of the theoretical maximum for the unsupported line. Such stubs take the form of a short length of short-circuited transmission line in shunt with the main transmission line. For frequencies at which the stub is an odd number of quarter wavelengths long, it presents zero admittance and therefore does not affect the main line. Obviously, it is often necessary to maintain a low stub admittance over a broad range of frequencies, and this may be affected by using an impedance transformer on the main line at the junction with the stub. The design of stubs and transformers can be found in the literature.¹²

Semiflexible Coaxial Lines

There are several commercially available coaxial lines having a limited permissible bending radius but exhibiting a loss only slightly greater than that of rigid coaxial lines. They are

characterized by a helical-, ribbed-, or foamed-dielectric support for a copper center conductor and a solid copper or aluminum tube for the outer conductor.¹³ A number of these have been qualified as MIL spec. components and are becoming more widely used. They are useful where their greater mechanical stability permits good phase matching and control in critical RF circuits.

51.6 FLEXIBLE COAXIAL LINES

Flexible coaxial line using a solid or stranded inner conductor, a plastic-dielectric support, and a braided outer conductor is probably the most common means of connecting together many separated components of a radio-frequency system. Although the transmission loss is relatively high because of the dielectric loss, the convenience often outweighs this factor in applications for which some loss is tolerable. Such lines are commercially available in a wide variety of size and impedance.

Table 51-3 summarizes the characteristics of the more common lines. The attenuation is shown in Figure 51-11, and the power-handling capacity in Figure 51-12.

Leakage of electrical energy through the braided outer conductor is sometimes a problem. The amount of leakage has been evaluated in two ways. One is to measure the equivalent coupling impedance to the outside environment, often another cable.^{14,15} Typical coupling impedance for a M17/74 cable has been reported as $150\ \mu\text{h}/\text{ft}$ from 100 to 5500 MHz. The other is to measure the shielding factor of the cable as the decibel loss through the cable-shielded braid.^{16,17} The coupling loss increases linearly with length (and the shielding decreases with length) and increases slightly at higher frequencies, where the openings between wires in the shield braid become a bigger part of a wavelength. Single-braid cable typically provides 30- to 40-dB shielding from 1 ft of cable. For longer lengths, shielding in decibels is reduced by $20\ \log$ (length in feet). Double-braid cable provides 60 to 80 dB. Special braid of flat metal strip provides 80 to 90 dB. Semirigid lines, which have a solid sheath, provide from 300 to several thousand decibels at the higher frequencies, where the skin depth becomes very small and the field inside penetrates to only a small part of the solid metal sheath.

Frequently used cable connectors are listed in Table 51-4.

51.7 HOLLOW-TUBE WAVEGUIDES

Electromagnetic energy can be propagated down hollow metal tubes, if the tubes are of sufficient size and are properly excited. The size of the tubes required usually limits their use to the very-high frequency (VHF) region or higher.

The energy can be propagated in a number of different types of waves as described next. In the usual case, waveguide devices are designed for transmission of a single wave type (most often the dominant wave or that having the lowest cutoff frequency) because the design problems are very greatly simplified.

Propagation Parameters

TE_{mn} waves: In the transverse electric waves, sometimes called H_{mn} waves, the electric vector is always perpendicular to the direction of propagation.

TM_{mn} waves: In the transverse magnetic waves, sometimes called the E_{mn} waves, the magnetic vector is always perpendicular to the direction of propagation.

TABLE 51-3 Flexible Coaxial Lines

M17 Number	Inner conductor, in (cm)	Dielectric	Diameter, dielectric, in (cm)	Number and type of shield braid	Jacket material	Outer diameter, in (cm)	Weight, lb/ft (Kg/m)	Nominal impedance, Ω	Nominal capacitance, pF/ft (pF/m)	Maximum temperature range, $^{\circ}\text{C}$	Maximum operating volts, rms	Comments
2-RG6	0.0285 CCS (0.0724)	PE	0.185 (0.469)	2: inner SC, outer BC 1: BC	PVC-IIA	0.332 (0.843)	0.082 (0.122)	75.0	20.6 (67.6)	-40 +85	3,000	Use to 3 GHz; NHL/180-00001
6-RG11	0.0477 (0.121) 7/0.0159 TC	PE	0.285 (0.724)	1: BC	PVC-IIA	0.405 (1.029)	0.098 (0.143)	75.0	20.6 (67.6)	-40 +80	5,000	Use to 1 GHz; NHL/181-00001; armored/6-RG12; high temp 62-RG144
15-RG22	2 \times 0.0456 (0.116) 7/0.0152 BC	PE	0.285 (0.724)	2: TC	PVC-IIA	0.420 (1.067)	0.134 (0.196)	95.0	16.0 (52.5)	-40 +85	1,000	Balanced; use to 200 MHz; NHL/182-00001; armored/15-RG111; NHL/182-00002
16-RG23	2 \times 0.0855 (0.217) 7/0.0285 BC	2 PE	0.380 (0.965)	2: separate inner com. Outer BC	PVC-IIA	0.945 (2.400)	0.530 (0.789)	125.0	12.0 (39.4)	-40 +85	7,000	Balanced; dual coaxial; use to 1 GHz; armored/16-RG24
19-RG25	0.0585 (0.149) 19/0.0117 TC	Rubber E	0.288 (0.732)	2: TC	Rubber-IV	0.505 (1.283)	0.205 (0.305)	48.0	50.0 (164.1)	-40 +90	10,000	Triaxial pulse; use to 1 MHz; armored/21-RG26
22-00001	0.0925 (0.235) 19/0.0185 TC	Rubber D	0.455 (1.156)	1: TC	Rubber-IV	0.595 (1.511)	0.300 (0.491)	48.0	55.0 (180.5)	-40 +85	15,000	Coaxial pulse; use to 1 MHz; armored/22-RG27
23-RG28	0.0925 (0.235) 19/0.0185 TC	Rubber D	0.455 (1.156)	2: TC, GS	Rubber-IV	0.735 (1.867)	0.400 (0.596)	48.0	50.0 (164.1)	-40 +85	15,000	Triaxial pulse; use to 1 MHz
24-RG34	0.0747 (0.190) 7/0.0249 BC	PE	0.460 (1.168)	1: BC	PVC-IIA	0.630 (1.600)	0.231 (0.344)	75.0	20.6 (67.6)	-40 +85	6,500	Use to 1 GHz

TABLE 51-3 Flexible Coaxial Lines (continued)

M17 Number	Inner conductor, in (cm)	Dielectric	Diameter, dielectric, in (cm)	Number and type of shield braid	Jacket material	Outer diameter, in (cm)	Weight, lb/ft (Kg / m)	Normal impedance, Ω	Normal capacitance, pF/ft (pF/m)	Maximum temperature range, $^{\circ}\text{C}$	Maximum operating voltage, rms	Comments
28-RG58	0.0355 (0.090) 19/0.0072 TC	PE	0.116 (0.295)	1: TC	PVC-IIA	0.195 (0.495)	0.026 (0.039)	50.0	30.8 (101.1)	-40 +85	1,900	SW; use to 1 GHz; NHL/183-00001; NS/155-00001; NHL/197-00001
29-RG59	0.0226 CCS (0.057)	PE	0.146 (0.371)	1: BC	PVC-IIA	0.242 (0.615)	0.035 (0.052)	75.0	20.6 (67.6)	-40 +85	2,300	Use to 1 GHz; NHL/184-00001
30-RG62	0.0253 CCS (0.064)	Air and PE	0.146 (0.371)	1: BC	PVC-IIA	0.242 (0.615)	0.038 (0.057)	93.0	13.5 (44.3)	-40 +80	1,000	Use to 1 GHz; NHL/185-00001
31-RG63	0.0253 CCS	Air and PE	0.285 (0.724)	1: BC	PVC-IIA	0.405 (1.029)	0.085 (0.131)	125.0	10.0 (32.8)	-40 +80	1,000	Use to 1 GHz; armored/31-RG79
33-RG64	0.0585 (0.149) 19/0.0117 TC	Rubber D	0.288 (0.782)	2: TC	Rubber-IV	0.460 (1.168)	0.220 (0.328)	48.0	55.0 (180.5)	-40 +85	10,000	Coaxial pulse; use to 1 GHz
34-RG65	0.0080 (0.020)	PE	0.285 (0.724)	1: BC	PVC-IIA	0.405 (1.029)	0.110 (0.164)	950	48.0 (157.5)	-40 +80	1,500	Coaxial time delay; use to 5 MHz
45-RG108	2 \times 0.0378 (0.096) 7/0.0126 TC	PE	0.079 (0.201)	1: TC	PVC-IIA	0.235 (0.597)	0.035 (0.052)	78.0	19.6 (64.3)	-40 +85	1,000	Balanced line; use to 10 MHz; NHL/186-00001
47-RG114	0.0070 CCS (0.018)	Air and PE	0.285 (0.724)	1: BC	PVC-IIA	0.405 (1.029)	0.089 (0.133)	185.0	6.5 (21.3)	-40 +80	1,000	Low capacitance; use to 1 GHz
52-RG119	0.1019 BC (0.259)	PTFE	0.332 (0.843)	2: BC	FG braid-V	0.465 (1.181)	0.228 (0.340)	50.0	29.4 (96.5)	-55 +200	6,000	SW; use to 3 GHz; NS/156-00001 armored/52-RG120

54-RG122	PE	0.0300 (0.078) 27/0.0050 TC	0.096 (0.244)	1: TC	PVC-IIA	0.160 (0.406)	0.021 (0.031)	50.0	30.8 (101.1)	-40 +85	1,900	SW; use to 1 GHz; NHL S/187-00001; NS/157-00001; NHL S/198-00001
56-RG130	PE	2 × 0.0855 BC (0.217) 7/0.0285	0.472 (1.199)	1: TC	PVC-IIA	0.625 (1.588)	0.300 (0.447)	95.0	16.3 (53.5)	-40 +85	3,000	Use to 200 MHz; armored/56-RG131
60-RG141	PTFE	0.0370 SCCS (0.099)	0.116 (0.295)	2: SC	FEP	0.195 (0.495)	0.043 (0.064)	50.0	29.4 (96.5)	-55 +200	1,900	SW; use to 8 GHz; NS/158-00001
64-RG164	PE	0.1045 BC (0.265)	0.680 (1.727)	1: BC	PVC-IIA	0.870 (2.210)	0.505 (0.752)	75.0	20.6 (67.6)	-40 +85	10,000	Use to 1 GHz; armored/64-RG35
64-RG165	PTFE	0.0960 (0.244) 7/0.0320 SC	0.285 (0.724)	1: SC	FG braid-V	0.410 (1.041)	0.121 (0.180)	50.0	29.4 (96.5)	-55 +250	2,500	SW; use to 3 GHz; armored/65-RG166; NS/159-00001
67-RG177	PE	0.195 BC (0.495)	0.680 (1.727)	2: SC	PVC-IIA	0.895 (2.273)	0.520 (0.774)	50.0	30.8 (101.1)	-40 +85	11,000	SW; use to 5.6 GHz; NS/160-00001
73-RG211	PTFE	0.1920 BC (0.4877)	0.620 (1.575)	1: BC	FG braid-V	0.730 (1.854)	0.516 (0.769)	50.0	29.4 (96.5)	-55 +250	7,000	SW; use to 1 GHz; NS/161-00002
73-RG212	PE	0.0556 SC (0.1412)	0.185 (0.470)	2: SC	PVC-IIA	0.332 (0.843)	0.089 (0.133)	50.0	30.8 (101.1)	-40 +85	3,000	SW; use to 10 GHz; NHL S/188-00001; NS/162-00001; NHL S/199-00001
74-RG213	PE	0.0888 (0.2256) 7/0.0296 BC	0.285 (0.724)	1: BC	PVC-IIA	0.405 (1.029)	0.111 (0.165)	50.0	30.8 (101.1)	-40 +85	5,000	SW; use to 1 GHz; armored/74-RG215; NHL S/189-00002; NS/163-00001; NHL S/189-00001

TABLE 51-3 Flexible Coaxial Lines (continued)

M17 Number	Inner conductor, in (cm)	Dielectric	Diameter, dielectric, in (cm)	Number and type of shield braid	Jacket material	Outer diameter, in (cm)	Weight, lb/ft (Kg/m)	Nominal impedance, Ω	Nominal capacitance, pF/ft (pF/m)	Maximum temperature range, $^{\circ}\text{C}$	Maximum operating volts, rms	Comments
75-RG214	0.0888 (0.2256) 7/0.0296 SC	PE	0.285 (0.724)	2: SC	PVC-IIA	0.425 (1.080)	0.130 (0.194)	50.0	30.8 (101.1)	-40 +85	5,000	SW; use to 10 GHz; low temp/75-RG365; NS/164-00002; NS/164-00001
77-RG214	0.0477 (0.1212) 7/0.0159 BC	PE	0.285 (0.724)	2: BC	PVC-IIA	0.425 (1.080)	0.124 (0.185)	75.0	20.6 (67.6)	-40 +85	5,000	Use to 3 GHz; NHL/191-00001
78-RG217	0.106 BC (0.269)	PE	0.370 (0.940)	2: BC	PVC-IIA	0.545 (1.384)	0.225 (0.299)	50.0	30.8 (101.1)	-40 +85	7,000	SW; use to 3 GHz; NS/165-00001; armored/165-00002; NHL/192-00001
79-RG218	0.195 BC (0.495)	PE	0.680 (1.727)	1: BC	PVC-IIA	0.870 (2.210)	0.510 (0.760)	50.0	30.8 (101.1)	-40 +85	11,000	SW; use to 1 GHz; NHL/193-00001; armored/79-RG219; NHL/193-00002
81-00001	0.260 BC 0.660	PE	0.910 (2.311)	1: BC	PVC-IIA	1.120 (2.845)	0.820 (1.221)	50.0	30.8 (101.1)	-40 +85	14,000	Use to 1 GHz; armored/81-00002
84-RG223	0.035 SC (0.089)	PE	0.116 (0.295)	2: SC	PVC-IIA	0.212 (0.539)	0.041 (0.061)	50.0	30.8 (101.1)	-40 +85	1,900	SW; use to 12.4 GHz; NHL/194-00001; NS/167-00001; NHL/200-00001
86-00001	0.0936 (0.238) 7/0.0312 SC	PTFE	0.285 (0.724)	2: SC	FG braid-V	0.430 (1.092)	0.195 (0.290)	50.0	29.4 (96.5)	-55 +200	5,000	Armored/86-00002
90-RG71	0.0253 CCS (0.064)	Air and PE	0.146 (0.371)	2: TC	PE-III A	0.245 (0.622)	0.050 (0.075)	93.0	13.5 (44.3)	-55 +85	1,000	Use to 1 GHz; NHL/195-00001

92-RG115	0.0840 (0.213)	Taped PTFE	0.255 (0.648)	2: SC	FG braid-V	0.415 (1.054)	0.185 (0.276)	50.0	29.4 (96.5)	-55 +200	5,000	SW; use to 12.4 GHz; w FEP jacket MS1792- 00001; NS/168-00001
93-RG178	0.0120 (0.030) 7/0.0040 SCCS	PTFE	0.033 (0.086)	1: SC	FEP-IX	0.071 (0.180)	0.0060 (0.0089)	50.0	29.4 (96.5)	-55 +200	1,000	SW; use to 3 GHz; high temp/93-00001, NS/169- 00001
94-RG179	0.0120 (0.0305) 7/0.0040 SCCS	PTFE	0.063 (0.160)	1: SC	FEP-IX	0.100 (0.254)	0.010 (0.015)	75.0	19.5 (64.0)	-55 +200	1,200	Use to 3 GHz; 95 Ω 95-RG180
95-RG210	0.0253 SCCS (0.0643)	Air and PTFE	0.146 (0.371)	1: SC	FG braid-V	0.242 (0.615)	0.050 (0.074)	93.0	13.5 (44.3)	-55 +200	1,000	Use to 3 GHz
100-RG133	0.0252 BC (0.064)	PE	0.285 (0.724)	1: BC	PVC-IIA	0.405 (0.020)	0.095 (0.142)	95.0	16.3 (53.5)	-40 +85	5,000	Use to 3 GHz
109-RG301	0.0009 (0.1547)	PTFE	0.185 (0.470)	1: HR	FEP-IX	0.245 (0.622)	0.056 0.083	50	29.4 96.5	-55 +200	3,000	Use to 3 GHz
110-RG302	0.0250 SCCS (0.064)	PTFE	0.146 (0.371)	1: SC	FEP-IX	0.206 (0.523)	0.031 (0.046)	75.0	19.5 (64.0)	-55 +200	2,300	Use to 3 GHz
111-RG303	0.0370 SCCS (0.094)	PTFE	0.116 (0.295)	1: SC	FEP-IX	0.170 (0.432)	0.031 (0.046)	50.0	29.4 (96.5)	-55 +200	1,900	SW; use to 3 GHz; NS/170-00001
112-RG304	0.0590 SCCS (0.150)	PTFE	0.185 (0.470)	2: SC	FEP-IX	0.280 (0.711)	0.094 (0.140)	50.0	29.4 (96.5)	-55 +200	3,000	SW; use to 8 GHz; NS/171-00001
113-RG316	0.0201 (0.051) 7/0.0067 SCCS	PTFE	0.060 (0.152)	1: SC	FEP-IX	0.098 (0.249)	0.012 (0.018)	50.0	29.4 (96.5)	-55 +200	1,200	SW; use to 3 GHz; NS/172-00001
116-RG307	0.0290 (0.074) 19/0.0058 SC	Foam PE	0.146 (0.371)	2: SC PUR inter-layer	PE-IIIA	0.265 (0.673)	0.080 (0.119)	75.0	16.9 (55.5)	-55 +80	1,000	Use to 1 GHz
119-RG174	0.0189 (0.048) 7/0.0063 CCS	PE	0.060 (0.152)	1: TC	PVC-IIA	0.110 (0.279)	0.009 (0.013)	50.0	30.8 (101.1)	-40 +85	1,500	SW; use to 1 GHz; NHLS/196-00001; NS/173-00001
124-RG328	0.4850 TC braid	Rubber H.J.H	1.065 (2.705)	3: TC, GS, TC	Neoprene	1.460 (3.708)	1.469 (2.186)	25.0	85.0 (278.9)	-55 +85	15,000	Use to 3 GHz

TABLE 51-3 Flexible Coaxial Lines (continued)

M17 Number	Inner conductor, in (cm)	Dielectric	Diameter, in (cm)	Number and type of shield braid	Jacket material	Outer diameter, in (cm)	Weight, lb/ft (Kg/m)	Nominal impedance, Ω	Nominal capacitance, pF/ft (pF/m)	Maximum temperature range, $^{\circ}\text{C}$	Maximum operating volts, rms	Comments
125-RG328	0.0585 (0.1486) 19/0.0117	TC rubber H.I.H	0.380 (0.980)	3: TC, TC	GS neo- prene	0.700 (1.778)	0.353 (0.524)	50.0	50.0 (164.1)	-10 +90	15,000	Use to 1 GHz
126-RG391	0.0477 (0.1212)	Cond PE & PE	0.295 (0.749)	1: TC	PUC-IIA	0.405 (1.029)	0.100 (0.148)	72.0	23.0 (75.5)	-40 +85	5,000	Use to 1 GHz; armored/126-RG392
127-RG393	0.094 SC (0.239) 7/0.312	PTFE	0.285 (0.724)	2: SC	FEP-IX	0.390 (0.991)	0.175 (0.261)	50.0	29.4 (96.5)	-55 +200	2,500	SW; use to 11 GHz; NS/174-00001
128-RG400	0.0384 SC (0.0975)	PTFE	0.116 (0.294)	2: SC	FEP-IX	0.195 (0.495)	0.050 (0.075)	50.0	29.4 (96.5)	-55 +200	1,900	SW; use to 12.4 GHz; NS/175-00001
131-RG403	0.012 SCCS (0.031) 7/0.004	PTFE	0.033 (0.0838)	2: SC FEP inter-layer	FEP-IX	0.116 (0.295)	0.015 (0.022)	50.0	29.4 (96.5)	-55 +200	1,000	SW; triaxial; use to 10 GHz
132-RG404	0.012 SCCS (0.031)	PTFE Cont. PTFE	0.033 (0.0838)	1: SC PE inter-layer	FEP-IX	0.071 (0.180)	0.018 (0.027)	50.0	30.4 (99.7)	-40 +200	1,000	Use to 1 GHz
134-00001	0.033 SC (0.083)	PE	0.116 (0.294)	2: SC PE inter-layer	PE-III A	0.245 (0.622)	0.045 (0.067)	50.0	30.8 (101.1)	-40 +70	1,900	SW; water-blocked; triaxial; use to 3 GHz; non WB M17/134-00002
135-00001	0.088 SC (0.224) 7/0.0296	PE	0.285 (0.723)	2: SC PE inter-layer	PUR	0.500 (1.270)	0.160 (0.237)	50.0	30.8 (101.1)	-40 +70	5,000	SW; water-blocked; triaxial; use to 3 GHz; non WB M17/135-00002
135-00003	0.081 SC (0.207)	PE	0.285 (0.728)	2: SC PE inter-layer	PE-III A	0.500 (1.270)	0.185 (0.274)	50.0	30.8 (101.1)	-40 +70	5,000	SW; water-blocked; triaxial; use to 3 GHz; non WB M17/135- 00004

136-00001	0.012 SCCS (0.030) 7/0.004	PTFE	0.063	1: SC	PFA-XIII	0.100 (0.254)	0.012 (0.0181)	75.0	19.5 (64.0)	-55 +230	1,200	Use to 3 GHz
137-00001	0.012 SCCS (0.030) 7/0.004	PTFE	0.102 (0.259)	1: SC	PFA-XIII	0.141 (0.358)	0.020 (0.0297)	95.0	15.4 (50.5)	-55 +200	1,500	Use to 3 GHz
138-00001	0.020 SCCS (0.051) 7/0.0067	PTFE	0.060 (0.152)	1: SC	PFA-XIII	0.098 (0.249)	0.012 (0.1810)	50.0	29.4 (96.5)	-55 +200	1,500	SW; use to 3 GHz
139-00001	0.012 SCBerC (0.030) 7/0.004	PTFE	0.102 (0.259)	1: SC CadBr	PFA-XIII	0.141 (0.358)	0.0178 (0.0264)	95.0	15.4 (50.5)	-55 +200	1,500	Use to 3 GHz
152-00001	0.020 SCCS (0.051) 7/0.0067	PTFE	0.060 (0.152)	2: SC	FEP-IX	0.114 (0.289)	0.0185 (0.274)	50.0	29.4 (96.5)	-55 +200	1,200	SW; use to 12.4 GHz
176-00002	2 cond. 0.0235 SCBerC (0.597) 19/0.005	PTFE	0.042 (0.107)	1: SCBerC	PFA-XIII	0.129 (0.328)	0.018 (0.267)	77.0	24.0 (78.8)	-55 +200	1,000	Use to 10 MHz
177-00001	0.012 SCCS (0.031) 7/0.004	PTFE	0.102 (0.259)	2: SC FEP inter-layer	FEP-IX	0.184 (0.467)	0.034 (0.057)	95.0	15.4 (50.5)	-55 +200	1,500	Use to 3 GHz
178-00001	0.012 SCCS (0.031) 7/0.004	PTFE	0.102 (0.259)	SC: NC; polyester braid NC CPC tape FEP inter- layer	SC: NC; polyester braid NC; FEP inter- layer	0.305 (0.775)	0.020 (0.029)	95.0	15.4 (50.5)	-55 +150	1,500	Use to 3 GHz
179-00001	0.012 SCCS (0.031) 7/0.004	PTFE	0.063 (0.160)	SC: NC; polyester CPC tape braid NC; FEP inter- layer	SC: NC; polyester CPC tape braid NC; FEP inter- layer	0.240 (0.610)	0.010 (0.015)	75.0	19.5 (64.0)	-55 +150	1,200	Use to 3 GHz

Abbreviations:
Dielectric
PE = solid polyethylene
PTFE = solid polytetrafluoroethylene

TABLE 51-3 Flexible Coaxial Lines (continued)

Abbreviations:	
PIB = polyisobutylene, Type B per MIL-C-17	
Rubber per MIL-C-170	
SIL = silicone rubber	
PS = polystyrene	
Conductors and braid materials	
BC = bare copper	
SC = silver-covered copper	
CCS = copper-covered steel	
TC = tinned copper	
SCCS = silver, copper-covered steel	
SCCAdBr = silver-covered cadmium bronze	
GS = galvanized steel	
TCCS = tinned copper-covered steel	
SSC = silver-covered strip	
CPC = copper polyester copper laminate	
NC = nickel-covered copper	
Jacket material	
PVC-I = black polyvinylchloride, contaminating, Type I, per MIL-C-17E	
PVC-II = gray polyvinylchloride, noncontaminating, Type II, per MIL-C-17E	
PVC-IIA = black polyvinylchloride, noncontaminating, Type IIA, per MIL-C-17Er	
PE-III = clear polyethylene	
PE-III A = high-molecular-weight black polyethylene, Type IIIA, per MIL-C-17E	
FG braid-V = fiberglass-impregnated, Type V, per MIL-C-17E	
FEP-IX = fluorinated ethylene propylene, Type IX, per MIL-C-17E	
PUR = polyurethane, black specific compounds	
SIL/DAC-VI = dacron braid over silicon rubber, Type VI, per MIL-C-17E Rubber per MIL-C-17E	
XLPE = cross-linked polyolefin	
Comments	

SW indicates spec. requirement for swept frequency attenuation and SWR testing.
NS/ indicates the M17 number for the unswept version of the same cable (recommended only to 400 MHz).
Armored/indicates the M17 number for the same cable with armor.
NHL S/indicates similar cable for nonhalogen, low-smoke environments (uses cross-linked polyolefin jacket).

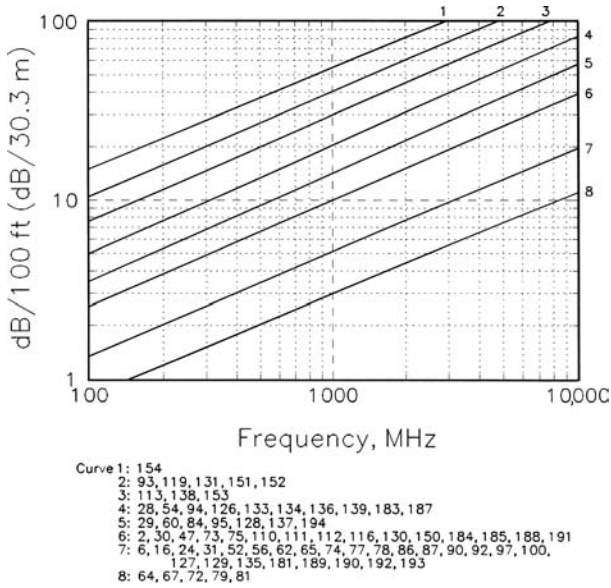


FIGURE 51-11 Approximate attenuation of flexible coaxial lines (M17 numbers are indicated on each curve.)

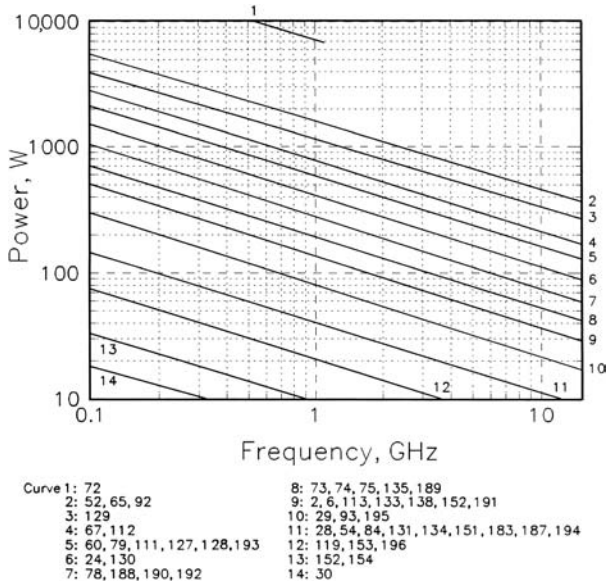


FIGURE 51-12 Approximate power-handling capacity of flexible coaxial lines. (M17 numbers are indicated for each curve.) Conditions are the following: ambient temperature 40°C, sea level, center conductor is 80°C for polyethylene and 200°C for PTFE.

TABLE 51-4 Characteristics of Frequently Used Cable Connectors

		BNC	TNC	C	SC	N
Cable size, outer diameters		in 0.150 –	0.150 –	0.300 –	0.300 –	0.330 – 0.550
		0.250 cm	0.250	0.550	0.550	0.762 – 1.40
		0.381 –	0.381 –	0.762 –	0.762 –	
		0.635	0.635	1.40	1.40	
Coupling type Bayonet (quick disconnect)		B		B		
Screw			7/16–28		11/16–24	5/8–24
Maximum operating volts		500	500	1000	1500	1000
Frequency range, GHz, DC to —		4	11	11	11	11
RF leakage, dB		–55	–60	–55	–90	–90
Insertion loss, dB		0.2	0.2	0.5	0.15	0.15
At GHz		3	3	10	10	10
Cable Types	Use	UG-	M39012/	UG-	M39012	UG-
M17/28, 111,	Plug (m)	88/U	26-0001			
	Plug, right-angle	913/U	30-0001			
60, 84	Cable jack (f)	89/U	27-0001			
	Panel jack (f)	262/U	29-0001		Flange nut	
		909/U	28-0001	704C/U		
M17/29 30,	Plug (m)	260/U	26-0002			
	Plug, right-angle		30-0002			
90, 110, 97	Cable jack (f)	261/U	27-0002			
	Panel jack (f)	291/U	29-0002		Flange nut	
		910/U	28-0002	631B/U		
	Panel jack (f)	290/U	32-0001	Flange nut		
		625/U	31-0001			
M17/17,* 73,	Plug (m)			626C/U	35-0001	18/U
	Plug, right-angle			710C/U	39-0001	
112	Cable jack (f)		Flange nut	633/U		20/U
	Panel jack (f)			629B/U		159B/U
				630B/U	40-0001	

TABLE 51-4 Characteristics of Frequently Used Cable Connectors (*continued*)

Cable Types	Use	UG-	M39012/ UG-	UG-	M39012	UG-
M17/74, 75, 77, 127	Plug (m)		Flange	573C/U	35-0002	21/U
	Cable jack (f)		nut	572B/U	40-0002	23/U
	Panel jack (f)			571B/U		160C/U
				570B/U		594/U
	Panel jack (f)		Flange	568A/U	41-0001	58/U
			nut	706B/U	42-0001	680/U
			nut	569B/U	43-0001	

NOTE: Underscored cables are not matched to connectors. m = male; f = female.

*Fits Types C and SC only.

† Fits Type N only.

The propagation constant γ_{mn} determines the amplitude and phase of each component of a wave as it is propagated along the waveguide. Each component may be represented by $A \exp(j\omega t - \gamma_{mn}z)$, where A is a constant, z is the distance along the direction of propagation, and $\omega = 2\pi f$. When γ_{mn} is real, there is no phase shift along the waveguide, but there is high attenuation. In fact, no propagation takes place, and the waveguide is considered below cutoff. The reactive attenuation L along the waveguide under these conditions is given by

$$L = \frac{54.58}{\lambda_c} \left[1 - \left(\frac{\lambda_c}{\lambda} \right)^2 \right]^{1/2} \text{ dB/unit length}$$

where λ is the wavelength in the unbounded medium and λ_c is the cutoff wavelength of that wave (a function of waveguide dimensions only). Waveguides are often used at frequencies far below cutoff as calibrated attenuators, since the rate of attenuation is determined by cross-section dimensions of the waveguide and the total attenuation in decibels is a linear function of the displacement of the output from the input.¹⁸

When γ_{mn} is imaginary, the amplitude of the wave remains constant but the phase changes with z and propagation takes place. γ_{mn} is a pure imaginary quantity only for lossless waveguide. In a practical case, γ_{mn} has both a real part α_{mn} , which is the attenuation constant, and an imaginary part β_{mn} , which is the phase constant; that is $\gamma_{mn} = \alpha_{mn} + j\beta_{mn}$.

The wavelength in a uniform waveguide is always greater than the wavelength in the unbounded medium and is given by

$$\lambda_g = \frac{\lambda}{[1 - (\lambda / \lambda_c)^2]^{1/2}}$$

The phase velocity is the apparent velocity, judging by the phase shift along the guide. Phase velocity, $v = c (\lambda_g / \lambda)$, is always greater than that in an unbounded medium.

The group velocity is the velocity of energy propagation down the guide. Group velocity, $u = c (\lambda / \lambda_g)$, is always less than that in an unbounded medium.

For air-filled guide and guides filled with dielectric having very low loss, the attenuation is mainly due to conductor losses in the walls. For any particular type of wave, the loss

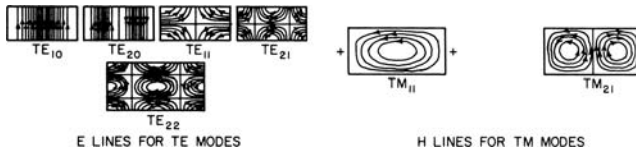


FIGURE 51-13 Field configurations for rectangular waveguide

is high near cutoff and decreases as the frequency is increased. For all excepting the TE_{0n} waves in circular waveguide, the attenuation reaches a minimum value for that wave and that waveguide, and then increases with frequency. For most waves, this minimum is slightly above $2f_c$. To avoid high loss near cutoff and the complexity of multiwave transmission, the useful band is usually considered to lie between $1.3f_c$ of the desired mode and $0.9f_c$ of the next higher mode. For rectangular waveguides having a width equal to twice the height, the useful range is about 1.5:1.

Rectangular Waveguides

For TE_{mn} waves in rectangular waveguides, m and n may take any integer value from 0 to infinity, except for the case $m = n = 0$. For the TM_{mn} waves, m and n may take any value from 1 to infinity. The m and n denote the number of the half-period variations of the electric field for TE waves or magnetic field for TM waves in the direction of the small and large dimensions respectively of the waveguide. Field patterns for some of the simpler waves are shown in Figure 51-13.

The propagation constant for rectangular guides is given by

$$\gamma_{mn} = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \omega^2\mu\epsilon} = \frac{2\pi}{\lambda_g}$$

where a is the wide dimension, b is the narrow dimension, ϵ is the dielectric constant, and μ is the permeability of the dielectric in the waveguide. Since propagation takes place only when the propagation constant is imaginary, the cutoff frequency for rectangular waveguide is

$$f_c = \frac{c}{2\sqrt{\mu\epsilon}} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}$$

and

$$\lambda_c = \frac{2\sqrt{\mu\epsilon}}{\sqrt{(m/a)^2 + (n/b)^2}}$$

Most frequently, operation is limited to the TE_{10} or dominant wave in rectangular waveguide. For this simplified case, the important formulas reduce to

$$\begin{aligned} \lambda_c &= 2a\sqrt{\mu\epsilon} \\ \gamma &= 2\pi\sqrt{(1/4a)^2 - f^2\mu\epsilon} \end{aligned}$$

In order to relate the waveguide properties to similar properties of low-frequency circuits, the impedance concept has been developed. Three characteristic impedances can be defined, differing from each other by a constant:¹⁹

$$Z_{VI} = \frac{V}{I} = 377 \frac{\pi}{2} \frac{b}{a} \frac{\lambda_g}{\lambda} = 592 \frac{b}{a} \frac{\lambda_g}{\lambda}$$

$$Z_{PV} = \frac{V^2}{2P} = 754 \frac{b}{a} \frac{\lambda_g}{\lambda} = 754 \frac{b}{a} \frac{\lambda_g}{\lambda}$$

$$Z_{PI} = \frac{2P}{I^2} = 377 \frac{\pi^2}{8} \frac{b}{a} \frac{\lambda_g}{\lambda} = 465 \frac{b}{a} \frac{\lambda_g}{\lambda}$$

Any one of the three is reasonably satisfactory if used consistently throughout, since the most frequent use is in determining mismatch at waveguide junctions and it is the ratio of impedance that matters. Ratios involving only different values of b give accurate indication of impedance mismatch. Differences in a give ratios nearly correct for small changes in a from the usual cross-section dimensions of rectangular waveguide, but errors are appreciable for large differences in a . Z_{PV} is most widely used, but Z_{VI} is found to be more nearly correct in matching coaxial line to waveguide.

Circular Waveguides

The usual coordinate system is ρ , θ , z , where ρ is the radial direction, θ is the angle, and z is the longitudinal direction.

For TE_{mn} waves in circular waveguides, m denotes the number of axial planes along which the normal component of electric field vanishes and n the number of cylinders including the boundary of the guide along which the tangential component of electric field vanishes. The number m may take any integral value from 0 to infinity, and n may take any integral value from 1 to infinity. The dominant wave in circular waveguide is the TE_{11} . For TM_{mn} waves, m denotes the number of axial planes along which the magnetic field vanishes and n the number of cylinders to which the electric field is normal. The number m may take any integral value from 0 to ∞ , and n may take any integral value from 1 to ∞ . Of the circularly symmetrical waves, the TM_{01} has the lowest cutoff frequency.

Field patterns for some of the simpler waves in circular guides are shown in Figure 51-14. The cutoff wavelength in lossless circular guide is given by

$$\lambda_c = \sqrt{\mu\epsilon} D_{mn} a$$

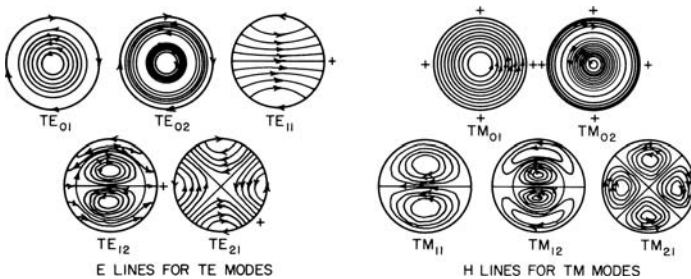


FIGURE 51-14 Field configurations for circular waveguide

where a is the radius, ϵ the dielectric constant, μ the permeability of the dielectric, and the constant D_{mn} is as shown in Table 51-5.

A summary of the cutoff and attenuation constant formulas for circular and rectangular waveguides is given in Table 51-6, where

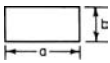
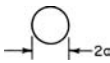
$$A = \frac{\sqrt{c/\lambda}}{\sqrt{1 - (\lambda/\lambda_c)^2}}$$

c = velocity of light in free space $\approx 3 \times 10^8$ m/s

TABLE 51-5 Cutoff Constants for Circular Waveguide

D_{mn} for TE _{<i>mn</i>} waves				
$m \backslash n$	0	1	2	3
1	1.640	3.412	2.057	1.496
2	0.896	1.178	0.937	0.764
3	0.618	0.736	0.631	0.554
	0.475	0.54	0.48	0.44
D_{mn} for TM _{<i>mn</i>} waves				
$m \backslash n$	0	1	2	3
1	2.613	1.640	1.224	0.966
2	1.139	0.896	0.747	0.644
3	0.726	0.618	0.541	0.482
4	0.534	0.475	0.425	0.388

TABLE 51-6 Summary of Cutoff and Attenuation Constant Formulas

TYPE OF GUIDE				
	RECTANGULAR TE _{<i>m0</i>}		TE _{<i>11</i>}	TE _{<i>01</i>}
CUTOFF WAVELENGTH	$\frac{2a}{m}$	$2.613a$	$3.412a$	$1.640a$
ATTENUATION CONSTANT (NEPERS/METER) DUE TO CONDUCTION LOSSES ONLY	$\frac{4\alpha_0 A}{a} \left(\frac{a}{2b} + \frac{\lambda^2}{\lambda_c^2} \right)$	$2 \frac{\alpha_0 A}{a}$	$2 \frac{\alpha_0}{a} A \left(0.415 + \frac{\lambda^2}{\lambda_c^2} \right)$	$2 \frac{\alpha_0}{a} A \frac{\lambda^2}{\lambda_c^2}$
ATTENUATION FAR BELOW CUTOFF	27.28 DB PER GUIDE WIDTH	41.78 DB PER DIAMETER	31.98 DB PER DIAMETER	66.56 DB PER DIAMETER

Lastly, a transition between a rectangular waveguide to a circular waveguide is a very useful component in antenna feed systems and one such device is described in the reference section.²⁰

For copper and air, $\alpha_0 = 3.5 \times 10^{-8}$ Np/m. To convert nepers per meter to decibels per 100 ft, multiply by 264.

Standard Waveguide Sizes

The waveguide sizes, which have become standardized, are listed in Table 51-7, together with the flanges used in connecting them together.

Flexible Waveguides

Flexible rectangular waveguides are made to match most of the standard waveguides. These differ primarily in mechanical construction. Waveguides using seamless corrugations, spiral-wound strip with adjacent edges crimped and soft-soldered, and spiral-wound strip with heavier crimping to provide sliding contact, as well as vertebra type (consisting of cover-choke wafers held in place by a rubber jacket), are available. The first two will bend in either plane, stretch, or compress but will not twist. The other two twist as well as bend and stretch. When the guide is flexed during operation or pressurized, it is nearly always covered with a molded-rubber jacket. When unjacketed, all are subject to a minimum bending radius (of the guide centerline) of 2 to 3 times the outer guide dimension in the plane of the bend, and when jacketed, to about 4 to 6 times the outer dimensions. The mismatch between rigid and flexible guide is small when straight and designed for lower frequencies. Mismatch increases as the waveguide size decreases, since the depth of convolutions cannot be decreased as fast as the waveguide dimensions (mismatch also increases as the bending radius is decreased). Similarly, attenuation which is only slightly greater than that in rigid guide for low frequencies becomes about twice as great at 40 GHz. Power capacity is nearly equal to that of rigid guide.

Hollow-Tube Waveguides with Other Cross Sections

One of the most useful of the many cross sections that might be used is the ridged waveguide as shown in Figure 51-15, which is useful in wide-band transmission. For ridged waveguide²¹⁻²³ (and waveguide of arbitrary cross section), the best method for obtaining cutoff wavelength is by resonance in the cross section. First, a convenient longitudinal plane is chosen. At cutoff, the susceptance looking into the shorted parallel-plate guide to the right of this plane is equal in amplitude and opposite in phase to that looking to the left. If the guide is symmetrical, only half of the guide need be used since Y at the center is zero. For the ridged guide this gives

$$\cot \frac{2\pi l}{\lambda_c} - \frac{B}{Y_{01}} = \frac{b}{a} \tan \frac{2\pi S}{\lambda_c 2}$$

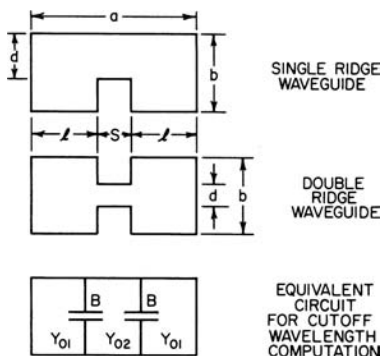


FIGURE 51-15 Ridged-waveguide configurations

where B is the capacitive discontinuity at the height change.

TABLE 51-7 Standard Rectangular Waveguides and Flanges

EIA designation WR-	DOD part no. M85/1-	AN designation	Material	Inside dimensions, in (mm)	Tolerance, in (mm)	Wall thickness, in (mm)	Frequency range, GHz for TE ₁₀	f_c , GHz for TE ₁₀	λ_c , mm	Attenuation dB/100 ft (dB/30.5 m)	Theoretical peak power, MW		Theoretical maximum continuous wave, kW	Flanges	
											Lowest and highest			Cover	Choke
650	017-	69/U	B	6.500 × 3.250 (165.1 × 82.55)	± 0.008 (0.20)	0.080 (2.03)	1.12–1.70	0.908	330.2	0.316–0.209 0.273–0.180	41.3–59.7	80.5–122 88.5–136	323/U 1720/U*	322/U	
	018-	103/U	A												
510	023-	337/U	B	5.100 × 2.550 (129.5 × 64.77)	± 0.008 (0.20)	0.080 (2.03)	1.45–2.20	1.154	259.1	0.440–0.299 0.380–0.258	26.2–37.0	47.9–70.4 53.2–78.3	1715/U* 1717/U*	1715/U*	
	025-	338/U	A												
430	029-	105/U	A	4.300 × 2.150 (109.2 × 54.61)	± 0.008 (0.20)	0.080 (2.03)	1.70–2.60	1.375	218.4	0.502–0.334 0.583–0.387	18.2–26.3	35.3–53.1 31.7–47.7	437/U* 435A/U*	437/U*	
	031-	104/U	B												
340	035-	113/U	A	3.400 × 1.700 (86.36 × 43.18)	± 0.006 (0.15)	0.080 (2.03)	2.20–3.30	1.737	172.7	0.682–0.474 0.791–0.550	11.9–16.4	21.7–31.3 19.5–28.1	554/U* 553A/U*	554/U*	
	037-	112/U	B												
284	041-	75/U	A	2.840 × 1.340 (72.14 × 34.04)	± 0.006 (0.15)	0.080 (2.03)	2.60–3.95	2.080	144.3	0.950–0.651 1.10–0.754	7.65–10.9	13.4–19.6 12.1–17.6	584/U 53/U	584/U	585A/U 54B/U
	043-	48/U	B												
229	047-	341/U	A	2.290 × 1.145 (58.16 × 29.08)	± 0.006 (0.15)	0.064 (1.63)	3.30–4.90	2.577	116.3	1.21–0.858 1.40–0.996	5.48–7.55	8.99–12.7 8.08–11.4	1727/U* 1726/U*	1727/U*	
	049-	340/U	B												
187	053-	95/U	A	1.872 × 0.872 (47.55 × 22.15)	± 0.005 (0.13)	0.064 (1.63)	3.95–5.85	3.155	95.10	1.79–1.24 2.07–1.44	3.30–4.70	5.17–7.45 4.64–6.69	407/U* 149A/U	407/U*	406B/U 148C/U
	055-	49/U	B												
159	059-	344/U	A	1.590 × 0.795 (40.39 × 20.19)	± 0.005 (0.13)	0.064 (1.63)	4.90–7.05	3.705	80.77	1.99–1.49 2.31–1.72	2.79–3.72	4.20–5.62 3.77–5.05	1731/U* 247/U	1731/U*	248/U 440B/U
	061-	343/U	B												
137	065-	106/U	A	1.372 × 0.622 (34.85 × 15.80)	± 0.004 (0.10)	0.064 (1.63)	5.85–8.20	4.285	69.70	2.53–2.00 2.94–2.32	1.98–2.53	2.90–3.67 2.60–3.29	441/U 344B/U	441/U	343B/U 137B/U
	067-	50/U	B												
112	071-	68/U	A	1.122 × 0.497	± 0.004	0.064	7.05–10.0	5.260	57.00	3.55–2.76	1.28–1.70	1.79–2.30	138/U		

90	073-	51/U	B	(28.50 × 12.62)	(0.10)	(1.63)	8.20–12.4	6.560	45.72	4.11–3.20	1.61–2.07	51/U	52B/U
	077-	67/U	A	0.900 × 0.400	± 0.004	0.050	10.0–15.0	7.847	38.10	5.54–3.83	0.959–1.39	135/U	136B/U
	079-	52/U	B	(22.86 × 10.16)	(0.10)	(1.27)	12.4–18.0	9.490	31.60	6.42–4.45	0.862–1.25	39/U	40B/U
75	083-	347/U	A	0.750 × 0.375	± 0.003	0.050	15.0–22.0	11.54	25.91	6.55–4.58	0.737–1.06		
	085-	346/U	B	(19.05 × 9.53)	(0.08)	(1.27)	18.0–26.5	14.08	21.38	7.60–5.31	0.662–0.948		
62	089-	91/U	B	0.622 × 0.311	± 0.0025	0.040	22.0–33.0	17.28	17.27	9.58–7.04	0.457–0.633	419/U	541A/U
	090-	349/U	A	(15.80 × 7.90)	(0.06)	(1.02)	26.5–40.0	21.10	14.22	8.26–6.07	0.502–0.683	1665/U	1666/U
51	093-	107/U	SC				40.0–60.0	30.69	9.550	6.91–5.08	0.602–0.818	419/U	541/U
	096-	353/U	B	0.510 × 0.255	± 0.0025	0.040	50.0–75.0	39.90	7.518	13.1–9.48	0.290–0.400		
	097-	351/U	A	(12.95 × 6.48)	(0.06)	(1.02)				11.3–8.71	0.323–0.445		
42	102-	53/U	B	0.420 × 0.170	± 0.002	0.040				20.05–15.0	0.157–0.213	595/U	596A/U
	103-	121/U	A	(10.67 × 4.32)	(0.05)	(1.02)				17.7–13.0	0.174–0.237	597/U	598A/U
34	106-	66/U	SC							14.8–10.9	0.209–0.284	595/U	596A/U
	109-	354/U	B	0.340 × 0.170	± 0.002	0.040				25.0–17.4	0.118–0.169	1530/U*	
	110-	355/U	A	(8.64 × 4.32)	(0.05)	(1.02)				21.6–15.0	0.131–0.188		
28	113-	357/U	SC				26.5–40.0	21.10	14.22	16.2–11.3	0.175–0.252	599/U	600A/U
	114-	96/U	S	0.280 × 0.140	± 0.0015	0.040				24.6–16.8	0.212–0.310		
22	117-	271/U	SC	(7.11 × 3.56)	(0.04)	(1.02)	33.0–50.0	26.35	11.38	22.0–15.1	0.242–0.347	383/U*	
	118-	97/U	S	0.224 × 0.112	± 0.0010	0.040				34.5–23.5	0.134–0.197		
19	121-	272/U	SC	(5.69 × 2.84)	(0.03)	(1.02)	40.0–60.0	30.69	9.550	31.0–21.1	0.149–0.219	1529/U*	
	124-	358/U	SC	0.188 × 0.094	± 0.0010	0.040				38.0–27.3	0.111–0.155		
15	125-	98/U	S	(4.78 × 2.39)	(0.03)	(1.02)					0.048–0.070		
	128-	273/U	SC	0.148 × 0.074	± 0.0010	0.040				64.2–43.9	0.052–0.077	385/U*	
				0(3.76 × 1.88)	(0.03)	(1.02)				57.6–39.3	0.058–0.085		

TABLE 51-7 Standard Rectangular Waveguides and Flanges (continued)

EIA designa- tion WR-	DOD part no. M85/1-	AN designa- tion	Material	Inside dimensions, in (mm)	Tolerance, in (mm)	Wall Thickness, in (mm)	Frequency range, GHz for TE ₁₀	$f_{c,TE_{10}}$ GHz for TE ₁₀	λ_c , mm	Attenuation dB/100 ft (dB/30.5 m)	Theoretical peak power, MW		Theoretical maximum continuous wave, kW	Flanges	
											Lowest and highest			Cover	Choke
12	129-	99/U	S	0.122 × 0.061 (3.10 × 1.55)	± 0.0005 (0.01)	0.040 (1.02)	60.0–90.0	48.40	6.198	87.8–58.9	0.020–0.030	0.047–0.067	0.047–0.067	387/U*	
10	132-	274/U	SC	0.100 × 0.050 (2.54 × 1.27)	± 0.0005 (0.01)	0.040 (1.02)	75.0–110	58.85	5.080	78.7–52.7	0.014–0.020	0.049–0.074	0.049–0.074	1528/U*	
	135-	359/U	SC												
8	138-	278/U	SC	0.080 × 0.040 (2.03 × 1.02)	± 0.0003 (0.01)	0.020 (1.51)	90.0–140	73.84	4.064	154–98.7	0.009–0.013	0.015–0.024	0.015–0.024	1527/U*	
7	141-	276/U	SC	0.065 × 0.0325 (1.65 × 0.83)	± 0.00025 (0.01)	0.020 (0.51)	110–170	90.85	3.302	214–135	0.006–0.009	0.010–0.016	0.010–0.016	1525/U*	
5	144-	275/U	SC	0.051 × 0.0255 (1.30 × 0.65)	± 0.00025 (0.01)	0.020 (0.51)	140–220	115.8	2.591	308–194	0.004–0.006	0.006–0.01	0.006–0.01	1524/U*	
4	147-	277/U	SC	0.043 × 0.0215 (1.09 × 0.546)	± 0.0002 (0.01)	0.020 (0.51)	170–260	137.5	2.184	377–251	0.003–0.005	0.005–0.007	0.005–0.007	1526/U*	
3	152-	139/U	S	0.034 × 0.0170 (0.864 × 0.432)	± 0.0002 (0.01)	Round	220–325	173.3	1.727	512–341	0.0004–0.0005	0.005–0.008	0.005–0.008		

Materials	Resistivity, $\mu\Omega \cdot \text{cm}$
A = aluminum alloy 1100	2.90
B = brass	3.90
SC = silver-clad copper	
S = silver	1.63

*These flanges mate with themselves.

The loading in the center of the guide lowers the cutoff frequency of the dominant mode, so that a useful bandwidth of over 4:1 may be obtained with single-mode transmission. The impedance is reduced by the loading and can be adjusted by proportioning the ridges for impedance matching of the waveguides to coaxial lines, for example.

The ridged guide for a given frequency band is smaller than the regular guide, but the losses are higher.

Design exists, and there are government standards for both single- and double-ridge guide with bandwidth ratios of 2:4 and 3:6. The loss of the 3:6-bandwidth guide approaches that of some of the low-loss coaxial lines and has not had wide acceptance. The 2:4-bandwidth double ridge is being used above 3.5 GHz, and the characteristics of the more popular sizes are listed in Table 51-8.

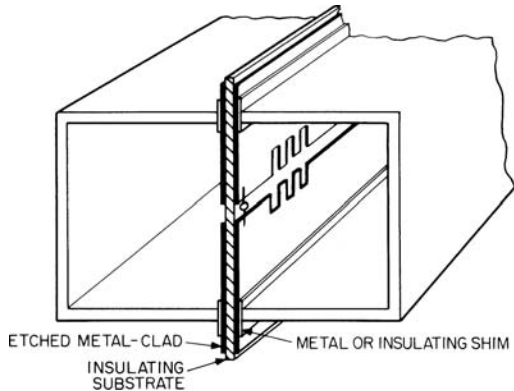


FIGURE 51-16 Fin-line guide

Fin-Line Guide Shown in Figure 51-16, fin-line guide uses central loading as in ridge guide, but the central-loading fin is an etched pattern on a suitable plastic or ceramic and may be insulated from the rest of the structure. It has the same wide bandwidth and concentration of the electric field at the guide center as the ridge guide, providing excellent excitation of the etched circuit patterns and any lumped circuit elements mounted on the center fin. Although not widely used as a uniform transmission line, fin-line guide is very useful when it is necessary to build filters, mixers, and other circuit elements right into the antenna feed.²⁴⁻²⁵ There are several different types of fin-lines, and design data for these fin-lines can be found in the references.²⁶⁻²⁷

51.8 MISCELLANEOUS TYPES OF WAVEGUIDES

In this section, the propagation parameters of the trough waveguide, radial line and biconical guide, dielectric waveguide, multimode waveguide, beam waveguide, optical-fiber waveguide, and elliptical waveguide are presented.

Trough Waveguide²⁸

Trough waveguide having the cross section shown in Figure 51-17 is derived from the lowest TE mode on symmetrical stripline by inserting a longitudinal electric wall at the center of the strip. Energy is transmitted in the TE modes in which the electric fields are symmetrical about the center vane. The cross section is such that probes, tuning, or attenuating devices may easily be inserted and moved from the open side with a minimum of electrical disturbance. Trough waveguide is a broadband transmission device since the cutoff of the second propagating mode is three times that of the dominant mode instead of twice, as in rectangular waveguide. It is easily fed from an end-on transition from coaxial or symmetrical stripline. The outer conductor of the TEM line is connected to the outer walls of the trough guide.

TABLE 51-8 Double-Ridge Waveguide: Bandwidth Ratio, 2.4:1

Type designation	DOD part no.	Material	Inside dimensions, in (mm)	Tolerance, in (mm)	Outside dimensions, in (mm)	Tolerance, in (mm)	Gap, in (mm)	Frequency range, GHz	Cutoff frequency, GHz		Theoretical attenuation, dB/100 ft (dB/ 30.5m)	Theoretical power, kW	Flange, UG-
									TE ₁₀	TE ₂₀			
350	029	A	1.480 × 0.688 (37.59 × 017.48)	± 0.003 (0.08)	1.608 × 0.816 (40.84 × 20.73)	± 0.004 (0.10)	0.2920 ± 0.002 7.417 (± 0.05)	3.50–8.20	2.915	8.620	3.07	151.3	1574/U
	030	B									3.03		1575/U
	031	C									2.04		1575/U
	032	S									2.18		1575/U
475	033	A	1.090 × 0.506 (27.69 × 12.85)	± 0.003 (0.08)	1.190 × 0.606 (30.23 × 15.39)	± 0.003 (0.08)	0.215 ± 0.002 (5.46) (± 0.05)	4.75–11.00	3.961	11.705	4.87	83.72	1577/U
	034	B									4.81		1578/U
	035	C									3.24		1578/U
	036	S									3.47		1578/U
750	037	A	0.691 × 0.321 (17.55 × 8.15)	± 0.003 (0.08)	0.791 × 0.421 (20.09 × 10.69)	± 0.003 (0.08)	0.136 ± 0.002 (3.45) (± 0.05)	7.50–18.00	6.239	18.464	9.64	33.58	1580/U
	038	B									9.51		1581/U
	039	C									6.41		1581/U
	040	S									6.86		1581/U
110	041	A	0.471 × 0.219 (11.96 × 5.56)	± 0.003 (0.08)	0.551 × 0.299 (14.00 × 7.59)	± 0.003 (0.08)	0.0930 ± 0.002	11.00–26.50	9.363	27.080	17.1	15.63	1583/U
	042	B									16.9		1584/U

Type designation	WRD no.	DOD part no. M23351/4-	Material	Inside dimensions, in (mm)	Tolerance, in (mm)	Outside dimensions, in (mm)	Tolerance, in (mm)	Gap, in (mm)	Frequency range, GHz	Cutoff frequency, GHz		Theoretical attenuation, dB/100 ft (dB/30.5m)	Theoretical power, kW	Flange, UG-
										TE ₁₀	TE ₂₀			
180		043	C					(2.362)				11.4		1584/U
		044	S					(± 0.05)				12.2		1584/U
		045	A	0.288 × 0.134	± 0.003	0.368 × 0.214	± 0.003	0.0570	18.00–40.00	14.495	44.285	35.8	5.834	1586/U
		046	B	(7.32 × 3.40)	(0.08)	(9.35 × 5.44)	(0.08)	±0.002				35.3		1587/U
		047	C					(1.448)				23.8		1587/U
		048	S					(± 0.05)				25.5		1587/U

Materials	Resistivity, $\mu\Omega \cdot \text{cm}$
A = aluminum alloy (6061)	4.0
B = brass	3.9
C = copper	1.77
S = coin silver	2.03

Note: Attenuation is computed at $f = 3f_{c10}$. Actual attenuation may be considerably higher, depending on operating frequency and temperature. Actual power handling may be considerably less, depending on temperature, altitude, operating frequency, and so on. See further information on both attenuation and power-handling capacity and on other design characteristics.²³

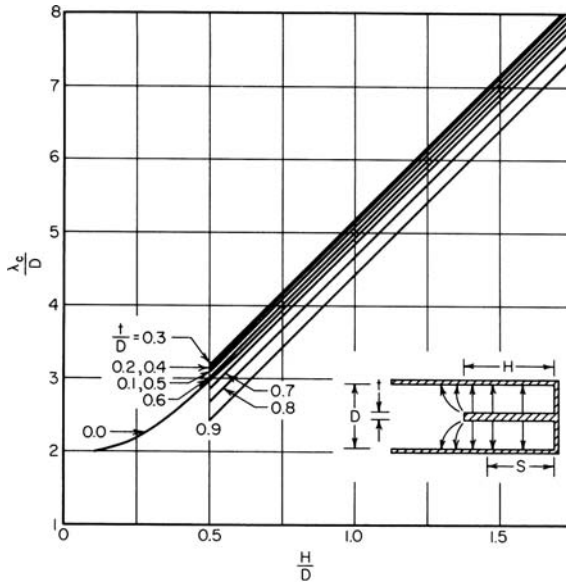


FIGURE 51-17 Cutoff wavelength of trough waveguide

The inner conductor of the line is connected at a point on the center vane of the trough guide at a distance S up from the bottom of the center vane, as determined from the relation for the power-voltage characteristic impedance at that point;

$$Z_{pV}(S) = 754 \frac{D}{\lambda_c} \frac{\lambda_g}{\lambda} \sin^2 \left(\frac{2\pi S}{\lambda_c} \right)$$

Alternatively, a coaxial connector can feed it from the rear wall of the trough.²⁹ This approach effectively shortens the length of the antenna structure and shields the coaxial transition from the radiation. Similarly, a crystal may be attached at an appropriate point to provide a crystal mount. The cutoff wavelength of the dominant mode in trough waveguide may be determined from the graph in Figure 51-17.

Radial Line and Biconical Guide

Two circular parallel conducting plates, separated by a dielectric and fed at the center or outer edge, form a line in which the transmission is radial. This type of line is frequently used in choke junctions and resonant cavities, such as microwave oscillator tubes. The simplest wave transmitted by this type of line is a TEM wave. The phase front of this wave is a circle of ever increasing or decreasing radius. The radial current in one plate returns radially through the other plate. With radial lines it is very useful to know the input impedance with (1) known termination, (2) output shorted, and (3) output open. Input impedance is

$$Z_i = Z_{oi} \left[\frac{Z_L \cos(\theta_i - \psi_L) + jZ_{0L} \sin(\theta_i - \theta_L)}{Z_{0L} \cos(\psi_i - \theta_L) + jZ_L \sin(\psi_i - \psi_L)} \right]$$

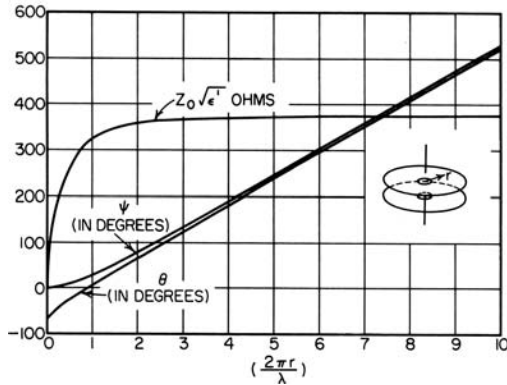


FIGURE 51-18 Radial transmission-line quantities

where

Z_i = input impedance, Ω

Z_{0i} = characteristic impedance, Ω , at input (Z_0 of Figure 51-18 at $r = r_i$)

Z_{0L} = characteristic impedance at output (Z_0 of Figure 51-18 at $r = r_L$)

Z_L = terminating impedance, Ω

$\theta_i, \theta_L, \psi_i, \psi_L$ = angles as plotted in Figure 51-18

In one special case of this, $Z_L = 0$; then

$$Z_i = jZ_{0i} \frac{\sin(\theta_i - \theta_L)}{\cos(\psi_i - \theta_L)}$$

In another $Z_L = \infty$; then

$$Z_i = jZ_{0i} \frac{\cos(\theta_i - \psi_L)}{\sin(\psi_i - \psi_L)}$$

Many higher-order modes are possible. Those with variations in ϕ will only propagate with any spacing of plates. Those having variations in z propagate only if the plate separation is greater than a half wavelength.^{1-2,30}

Two cones with their apices facing and fed by a balanced input at the center, as shown in Figure 51-19, form a biconical waveguide. This structure also simulates a dipole antenna and certain classes of cavity resonators.

One important wave transmitted by this type of guide has no radial components and propagates with the velocity of light along the cones. It is analogous to the TEM wave in cylindrical systems. The ratio of voltage to current or characteristic impedance is

$$Z_0 = 120 \log_e \cot \frac{\theta}{2} \quad \Omega$$

where θ is the conical angle. From this it is seen that the characteristic impedance is independent of radius and not variable as in the radial line.

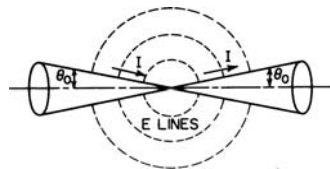


FIGURE 51-19 Biconical waveguide

Many higher modes can propagate on this system. These will all propagate at a velocity different from that of light. Transmission of higher modes is also possible.¹⁻²

Dielectric Waveguide

Electromagnetic waves will propagate along a dielectric rod,³¹ if the rod is of sufficient size. At low frequencies, there is little advantage to dielectric wave-

guides since low loss in metal waveguide makes them a convenient shielded carrier of microwave energy. At frequencies above 20 GHz, the lower loss possible with dielectric waveguides makes their use attractive. For lossless dielectric waveguides, the propagation constant can never be real, so there is no cutoff frequency as with metal waveguides. For a given waveguide, VHF energy is confined entirely within the dielectric. The velocity of propagation and the loss correspond to that in the waveguide dielectric. As the frequency is reduced, more of the field is outside the waveguide, and the velocity and loss approach that of the surrounding air, but the dielectric ceases to guide the wave. The only TE and TM modes possible in a circular rod are those having axial symmetry. One hybrid mode, the HE_{11} (hybrid because it has both E_z and H_z components), is particularly well adapted for microwave transmission. It can be small (it is the only mode which can be propagated when the ratio of diameter to wavelength is low; less than 0.6 for polystyrene). It can be launched from the dominant TE metallic waveguide mode, and it has low loss. The approximate field configuration of this mode is shown in Figure 51-20.

Figures 51-21 to 51-23 show variation of guide wavelength with diameter, loss as a function of diameter, and a waveguide launcher for the HE_{11} hybrid mode. The polarization of the hybrid mode in circular rod is subject to rotation because of internal stresses, dimensional nonuniformity, and bends. A rectangular or oval cross section prevents this depolarization. Measurements of loss and radius of field extent (radius at which field decreases to $1/e$ times that at surface) for cross sections of the oval type are shown in Table 51-9 for 24 and 48 GHz.

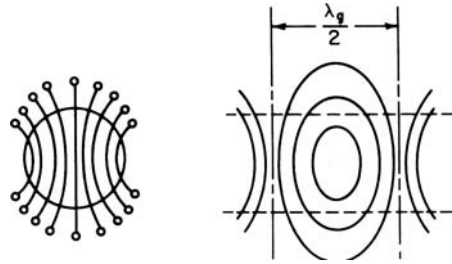


FIGURE 51-20 Approximate E -field configuration of HE_{11} mode on a dielectric rod waveguide

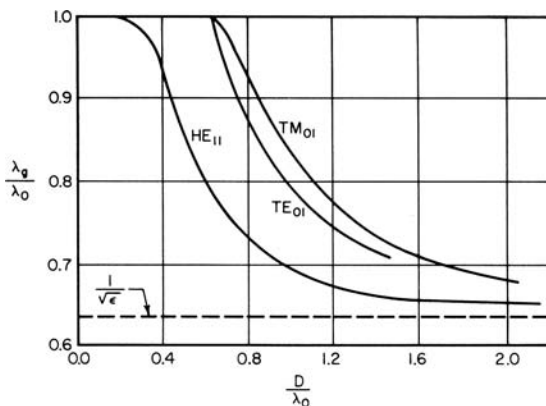







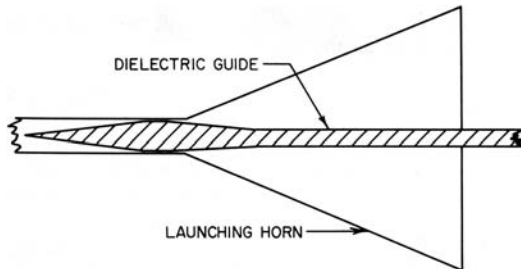
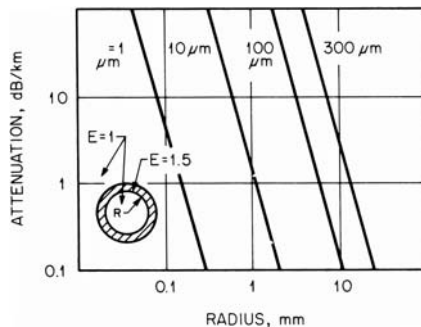


FIGURE 51-21 Guide wavelength versus diameter for polystyrene waveguide

TABLE 51-9 Loss for HE_{11} Mode in Dielectric Waveguide

Material Polystyrene	Dimensions in (cm)	E Direction	Radius of Field Extent, in (cm)	Loss dB/km	Freq. GHz
SHEET STOCK	.095 × .156 (.241 × .369)		1.5 (3.8)	164	24
EXTRUDED	.086 × .155 (.218 × .394)		4 (10.2)	164	
EXTRUDED	.086 × .155 (.218 × .394)		7 (17.8)	82	
EXTRUDED	.056 × .142 (.142 × .361)		0.4 (1.0)	3281	48
EXTRUDED	.056 × .142 (.142 × .361)		0.8 (2.0)	886	
EXTRUDED	.038 × .114 (.097 × .290)		1.5 (3.8)	148	
EXTRUDED	.032 × .096 (.081 × .244)		3 (7.6)	23	

**FIGURE 51-22** Waveguide launcher for HE_{11} mode**FIGURE 51-23** Attenuation of dielectric tube wave-length with a dielectric of 1.5 for the HE_{11} mode

Transmission with dielectric tubes as well as rods is possible. Tubes can have lower loss than rods, and theoretical calculations indicate³² that with a polystyrene tube attenuation at 30 GHz compares with TE_{01} guide.

Other calculations shown in Figure 51-23 indicate that losses of a few decibels per kilometer can be attained up to the infrared region by choice of the proper diameter.³³⁻³⁴

The image line³⁵⁻³⁷ is a variation of the dielectric guide which simplifies the support problem. The HE_{11} mode, being symmetrical, can be split longitudinally and one-half replaced by a metal plate, as shown in Figure 51-24. Losses at bends are lower for concave bends than for convex ones. Losses in the image plane are lower than those in the dielectric for all commonly used materials.

A variation of the image line has been suggested for use in integrated circuits. In this version,³⁸ a rectangular dielectric strip is laid on top of an integrated circuit board, which is copper-clad on the lower surface. An overview on dielectric waveguide microwave integrated circuits technology and system applications can be found in the reference section.³⁹ Tapering the dielectric strip and inserting it into a rectangular waveguide normally excite the dominant mode in a dielectric waveguide. However, a microstrip line can also excite dielectric waveguides and one such transition was developed at about 15 GHz.⁴⁰

A disadvantage of an open dielectric waveguide is its high radiation loss, which occurs at bends and other discontinuities. This spurious radiation not only increases the attenuation but also causes interference with other parts of the circuit. To overcome this problem, a non-radiative dielectric (NRD) waveguide has been proposed.⁴¹ In this structure, the dielectric waveguide is sandwiched between two metal plates separated by a distance smaller than half wavelength.

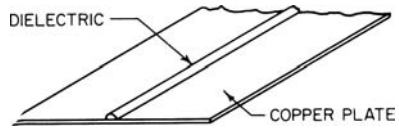


FIGURE 51-24 Image line using HE_{11} mode

Multimode Waveguide

Normally, waveguides are used in the lowest mode, and the higher modes are avoided. At millimeter-wave frequencies, losses in the lowest-mode-only waveguide become very high because of conductor losses. By using low-loss modes in the circular waveguide far above cutoff, most of the energy is carried in the fields in the air inside and losses are reduced. Operating far from cutoff also reduces the dispersion and increases the signal bandwidth. Special efforts must be taken to prevent excitation of unwanted modes. Multimode circular guide has demonstrated a signal bandwidth of 40 to 117 GHz and transmission losses of less than 1 dB/km.⁴²

Beam Waveguide

Another method of avoiding conductor loss is to use lenses to focus the energy into a converging beam or at least a nondiverging beam, refocusing periodically along the transmission path.⁴³

Optical-Fiber Waveguide

Optical fiber is usually a single strand of glass designed to hold an electromagnetic wave closely bound to the glass signal path. Since it is not a carrier of the fundamental signal frequency (a microwave signal would ordinarily be modulated on a light beam), it is not covered in detail here. It has many of the advantages of the multimode waveguide, the beam waveguide, and the

dielectric waveguide, principally low loss and large bandwidth. A small amount of dispersion limits the distance that a wideband signal can go before some frequencies overtake others and cause information degradation. However, operational use has demonstrated losses well under 1 dB/km and bandwidths of over 1 GHz divided by the distance in kilometers.⁴⁴ It will undoubtedly become a major transmission medium for wideband signals.

Semiconductor laser diodes and photo detectors with gigahertz bandwidth have been recently demonstrated. These devices allow the optical carrier to be directly modulated at gigahertz frequencies⁴⁵ and hence have given rise to a new field of optically controlled microwave devices.⁴⁶ This new field has enabled optically controlled beam scanning of active phased arrays and true time delays phase shifters for phased arrays.⁴⁷

Elliptical Waveguide

Waveguide with an elliptical cross section is very useful for antenna feeds. It has many of the same characteristics as rectangular waveguide, including a similar waveguide mode (so it mates easily with rectangular guide), and about the same loss and guide wavelength. It does have a narrower bandwidth, and it is formed with corrugations along its length, which gives it high transverse stability and crust strength yet allows bending in both planes and a small amount of twist. This makes installation easier than with rectangular guide. It is usually covered with tough black polyethylene to protect it during the hanging, installation, and use. Elliptical waveguide is made in long lengths or is assembled in desired sections with an adapting flange to rectangular guide. Although not yet standardized, it is made by several manufacturers. Most give it a prefix indicating elliptical waveguide such as EW or WE, followed by the lowest frequency or a midfrequency in tenths of a gigahertz. Characteristics of typical sizes are shown in Table 51-10.

51.9 LIST OF SYMBOLS

The symbols used in this chapter are listed here:

f = frequency, Hz

$\omega = 2\pi f$

c = velocity of light $\approx 3 \times 10^8$ m/s

λ = wavelength

ϵ = relative dielectric constant

ϵ_0 = permittivity of free space, 8.854×10^{-12} F/m

μ = relative permeability

μ_0 = permeability of free space, $4\pi \times 10^{-7}$ H/m

σ = conductivity, S/m

α = attenuation constant, Np/m (Np = 8.686 dB)

E_a = breakdown of air, V/m

P = power, W

$\tan \delta$ = loss tangent or dissipation factor

The properties of some commonly used metals will be found in Chapter 55. The conductivity σ can be found from the resistivity values by the relationship

$$\sigma(\text{S/m}) = \frac{10^8}{\text{resistivity } (\mu\Omega \cdot \text{cm})}$$

Values for ϵ and $\tan \delta$ of some commonly used dielectrics will be found in Chapter 55.

TABLE 51-10 Typical Characteristics of Elliptical Waveguide

Type	Maximum potential operating range, GHz	TE ₁₁ - mode cutoff frequency, GHz	Major and minor dimensions over jacket, in (mm)	Bending-radii minimum, in (mm)		Recommended twist, °/ft (m)
				<i>E</i> plane	<i>H</i> plane	
EW20	1.9–2.7	1.60	5.02×2.83 (127.5 \times 71.9)	26 (660)	71 (1800)	0.25 (0.75)
EW28	2.6–3.5	2.20	3.65×2.33 (92.5 \times 59.2)	22 (560)	52 (1320)	0.25 (0.75)
EW37	3.3–4.3	2.81	2.90×1.86 (73.7 \times 47.2)	17 (430)	41 (1040)	0.5 (1.5)
WE37	3.4–4.2	2.81	3.04×1.93 (71.0 \times 49.0)	19.7 (500)	39.4 (1000)	0.5 (1.5)
EW44	4.2–5.1	3.58	2.31×1.59 (58.7 \times 40.4)	15 (380)	32 (810)	0.5 (1.5)
WE44	4.2–5.0	3.58	2.24×1.59 (520 \times 43.0)	15.7 (400)	47.2 (1200)	0.5 (1.5)
EW52	4.6–6.425	3.63	2.25×1.31 (57.2 \times 33.3)	12 (305)	32 (810)	1 (3)
WE56	5.4–6.425	4.08	1.97×1.30 (50.0 \times 33.0)	19.7 (500)	39.4 (1000)	1 (3)
WE59	5.9–7.15	4.34	1.85×1.22 (47.0 \times 31.0)	11.8 (300)	23.5 (600)	1 (3)
EW63	5.85–7.125	3.96	2.01×1.16 (51.1 \times 29.5)	10 (260)	29 (740)	1 (3)
EW64	5.3–7.75	4.36	1.91×1.12 (48.5 \times 28.4)	10 (260)	27 (685)	1 (3)
WE64	6.4–7.75	4.16	1.84×1.14 (49.0 \times 29.0)	11.8 (300)	23.6 (600)	1 (3)
WE71	7.1–8.6	5.23	1.54×0.98 (390 \times 25.0)	11.8 (300)	23.6 (600)	1 (3)
EW77	6.1–8.5	4.72	1.72×1.00 (43.6 \times 25.4)	9 (230)	25 (635)	1 (3)
EW85	7.7–10.0	6.55	1.32×0.90 (33.5 \times 22.9)	8 (200)	19 (480)	1 (3)
WE85	8.5–10.2	5.83	1.38×0.94 (35.0 \times 24.0)	88 (200)	16 (400)	1 (3)
EW90	8.3–11.7	6.50	1.32×0.80 (33.5 \times 20.3)	7 (180)	19 (480)	2 (6)
WE107	8.5–11.7	6.80	1.18×0.79 (30.0 \times 20.0)	6 (150)	16 (400)	2 (6)
WE122	12.2–13.25	7.29	1.09×0.71 (28.0 \times 18.0)	6 (150)	16 (400)	2 (6)
EW122	10.0–13.25	8.46	1.07×0.72 (27.2 \times 18.3)	6 (150)	15 (380)	2 (6)
EW132	11.0–15.35	9.33	0.96×0.61 (24.4 \times 15.5)	5 (130)	14 (360)	2 (6)

REFERENCES

1. S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3rd Ed. (New York: John Wiley & Sons, Inc., 1994).
2. C. G. Montgomery, R. H. Dicke, and E. M. Purcell (eds.), *Principles of Microwave Circuits* (New York: McGraw-Hill Book Company, 1948).
3. R. M. Chisholm, "The Characteristic Impedance of Trough and Slab Lines," *IRE Trans. Microwave Theory Tech.*, vol. MTT-4 (July 1956): 166–172.
4. K. C. Gupta, R. Garg, I. J. Bahl, and P. Bhartia, *Microstrip Lines and Slot Lines* (Dedham, MA: Artech House, Inc., 1996).
5. B. C. Wadell, *Transmission Line Design Handbook* (Norwood, MA: Artech House Inc., 1991).
6. R. N. Simons, *Coplanar Waveguide Circuits, Components, and Systems* (New York: John Wiley & Sons, Inc., 2001).
7. R. Janaswamy and D. H. Schaubert, "Characteristic Impedance of a Wide Slotline on Low-Permittivity Substrates," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34 (August 1986): 900–902.
8. R. Q. Lee and R. N. Simons, *Advances in Microstrip and Printed Antennas*, Chap. 9, K. F. Lee and W. Chen (eds.) (New York: John Wiley & Sons, Inc., 1997): 443–514.
9. A. F. Hinte, G. V. Kopcsay, and J. J. Taub, "Choosing a Transmission Line," *Microwaves*, December 1971: 46–50.
10. J. M. Schellenberg, "CAD Models for Suspended and Inverted Microstrip," *IEEE Trans. Microwave Theory Tech.*, vol. 43 (June 1995): 1247–1252.
11. W. Sichak, "Coaxial Line with Helical Conductor," *IRE Proc.*, vol. 42 (August 1954): 1315–1319; also correction, *IRE Proc.*, vol. 43 (February 1955): 148.
12. G. L. Ragan (ed.), *Microwave Transmission Circuits* (New York: McGraw-Hill Book Company, 1948).
13. J. W. E. Griensmann, "An Approximate Analysis of Coaxial Line with Helical Dielectric Support," *IRE Trans. Microwave Theory Tech.*, vol. MTT-4 (January 1956): 13–23.
14. Greenblatt, J. W. E. Griensmann, and L. Birenbaum, "Measurement of Energy Leakage from Cables at VHF and Microwave Frequencies," AIEE Conf. pap., winter general meeting, 1956.
15. A. A. Smith, Jr., *Coupling of External Electromagnetic Fields to Transmission Lines* (New York: John Wiley & Sons, Inc., 1977).
16. H. W. Ott, *Noise Reduction Techniques in Electronic Systems* (New York: John Wiley & Sons, Inc., 1988).
17. *Coaxial Cable: The Complete Catalog & Handbook* (Wallingford, CN: Times Wire and Cable Company, 1987).
18. H. A. Wheeler, "The Piston Attenuator in a Waveguide Below Cutoff," *Wheeler Monog.*, vol. 8, (Great Neck, NY: Wheeler Laboratories, Inc., 1949).
19. S. A. Schelkunoff, *Electromagnetic Waves* (Princeton, N.J.: D. Van Nostrand Company, Inc., 1943): 319.
20. E. L. Holzman, "A Simple Circular-to-Rectangular Waveguide Transition," *IEEE Microwave & Wireless Components Letts.*, vol. 15 (January 2005): 25–26.
21. Radio Research Laboratory staff, *Very High Frequency Techniques* (New York: McGraw-Hill Book Company, 1947).
22. S. Hopfer, "The Design of Ridged Waveguides," *IRE Trans. Microwave Theory Tech.*, vol. MTT-3, (October 1955): 20–29.
23. Tsung-Shan Chen, "Calculation of the Parameters of Ridge Waveguides," *IRE Trans. Microwave Theory Tech.*, vol. MTT-5 (January 1957): 12–17.
24. P. J. Meier, "Integrated Fin-Line Millimeter Components," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-22 (December 1974): 1209–1216.

25. P. J. Meier, "Millimeter Integrated Circuits Suspended in the E-Plane of Rectangular Waveguide," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26 (October 1978): 726–732.
26. J. B. Knorr and P. M. Shayda, "Millimeter-Wave Fin-Line Characteristics," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (July 1980): 737–743.
27. R. N. Simons, "Analysis of Millimeter-Wave Integrated Fin Line," *IEE Proc.*, vol. 130, Pt. H (March 1983): 166–169.
28. K. S. Packard, "The Cutoff Wavelength of Trough Waveguide," *IRE Trans. Microwave Theory Tech.*, vol. MTT-6, no. 4 (October 1958): 455–456.
29. G. D. Hopkins, B. K. Edenfield, J. G. Hampton, and R. L. Roberts, "A New Coax to Troughguide Transition," *IEEE Microwave & Wireless Components Letts.*, vol. 12 (August 2002): 299–301.
30. N. Marcuvitz (ed.), *Waveguide Handbook* (London, U.K.: Institute of Electrical Engineers, 1986).
31. M. T. Weiss and E. M. Gyorgy, "Low Loss Dielectric Waveguides," *IRE Trans. Microwave Theory Tech.*, vol. MTT-2 (September 1954): 38–44.
32. *Arch. Elek. Übertragung*, vol. 8, no. 6 (June 1954).
33. M. Miyagi and S. Nishida, "A Proposal of Low-Loss Waveguide for Submillimeter Wave Transmission," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (April 1980).
34. M. Miyagi and S. Nishida, "Transmission Characteristics of Dielectric Tube Leaky Waveguide," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-28 (June 1980).
35. D. D. King, "Circuit Components in Dielectric Image Lines," *IRE Trans. Microwave Theory Tech.*, vol. MTT-3 (December 1955): 35–39.
36. D. D. King, "Properties of Dielectric Image Lines," *IRE Trans. Microwave Theory Tech.*, vol. MTT-3 (March 1955): 75–81.
37. S. Shindo and Itanami, "Low Loss Rectangular Dielectric Image Line for Millimeter-Wave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26 (October 1978): 747–751.
38. W. V. McLevige, T. Itoh, and R. Mittra, "New Waveguide Structures for Millimeter-Wave and Optical Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23 (October 1975): 788–794.
39. R. M. Knox, "Dielectric Waveguide Microwave Integrated Circuits—An Overview," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-24 (November 1976): 806–814.
40. H. Tehrani, M.-Y. Li, and K. Chang, "Broadband Microstrip to Dielectric Image Line Transition," *IEEE Microwave & Guided Wave Letts.*, vol. 10 (October 2000): 409–411.
41. T. Yoneyama and S. Nishida, "Nonradiative Dielectric waveguide for Millimeter-Wave Integrated Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29 (November 1981): 1188–1192.
42. "Millimeter Waveguide Systems," *Microwave J.* (March 1977): 24–26.
43. J. A. Arnaud and J. T. Ruscio, "Guidance of 100 GHz Beams by Cylindrical Mirrors," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23 (April 1975): 377–379.
44. K. Shirahata, W. Susaki, and H. Mamizaki, "Recent Developments in Fiber Optic Devices," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-30 (February 1982): 121–130.
45. K. Iizuka, *Elements of Photonics, Vol. II, For Fiber and Integrated Optics* (New York: John Wiley & Sons, Inc., 2002).
46. R. N. Simons, *Optical Control of Microwave Devices* (Norwood, MA: Artech House Inc., 1990).
47. A. Kumar, *Antenna Design with Fiber Optics* (Norwood, MA: Artech House Inc., 1996).

Chapter 52

Impedance Matching, Broadbanding, and Baluns

David F. Bowman

Antenna Consultant

Edward F. Kuester

University of Colorado at Boulder

CONTENTS

52.1	GENERAL	52-2
52.2	IMPEDANCE MATCHING WITH LUMPED ELEMENTS	52-5
52.3	IMPEDANCE MATCHING WITH DISTRIBUTED ELEMENTS	52-10
52.4	COMBINATIONS OF TRANSFORMERS AND STUBS	52-18
52.5	TAPERED LINES	52-20
52.6	BROADBANDING	52-22
52.7	BALUNS	52-26

52.1 GENERAL¹

Impedance matching is the control of impedance for the purpose of obtaining maximum power transfer or minimum reflection. This chapter describes circuits and techniques used for impedance matching, with emphasis on those most suitable for broadband operation. The impedance-matching methods of this chapter are limited to the use of linear, passive, and reciprocal elements.

Impedance Matching for Maximum Power Transfer

A load impedance connected to a source will receive the maximum possible power (the *available power* P_0) from the source when its value is adjusted to equal the complex conjugate of the impedance of the source (Figure 52-1a). The load impedance and source impedance are then matched on a conjugate-impedance basis (or simply *conjugate-matched*).

A mismatched load impedance $Z_L \neq Z_G^*$ will receive a reduced power $P = P_0 (1 - |m|^2)$, where m is the conjugate mismatch reflection coefficient

$$m = \frac{Z_L - Z_G^*}{Z_L + Z_G}$$

The mismatch may be eliminated by a properly designed network (Figure 52-1b) interposed between them. This “matching” network transforms the source impedance to the

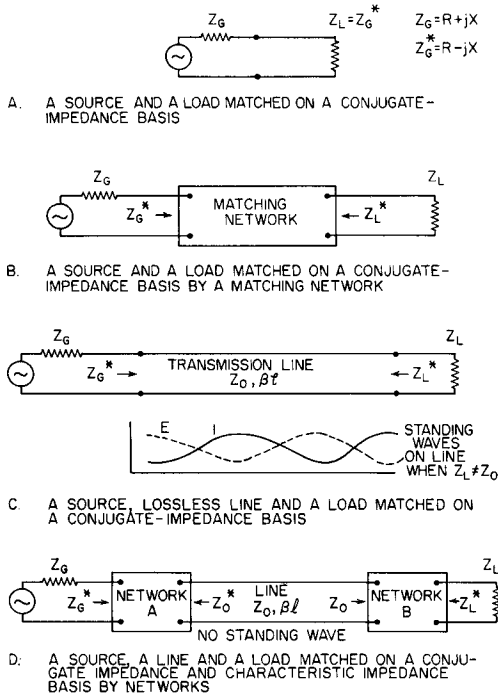


FIGURE 52-1 Impedance matching of transmission circuits

conjugate of the load impedance, and vice versa. Thus a conjugate-impedance match occurs at the input to the network and at the output of the network. Since maximum power transfer is the goal, it is usually the case that the matching network is lossless. It is true of any lossless transmission circuit that if a conjugate-impedance match is obtained at any point along the transmission path, then a conjugate-impedance match is obtained at all other points along the path.

Impedance Matching for Minimum Reflection

A length of transmission line may form one link in a transmission circuit. The line in general supports two waves (direct and reflected), propagating in opposite directions. In passing from one end of a length ℓ of line to the other, each wave is phase-delayed by an amount $\beta\ell$ (called the electrical length of the line) in its direction of propagation, where β is the phase constant or propagation constant of the line measured in units of rad/m. The interference between these waves creates standing-wave patterns of voltage and current and time-delayed “ghosts” of the direct signal. Sometimes these effects are tolerable, but usually they are undesirable. The voltage reflection coefficient at a load junction is the ratio of phasor voltage in the reflected wave to that in the direct or incident wave, and is given by

$$\rho = |\rho| \exp(j\phi) = \frac{Z_L - Z_0}{Z_L + Z_0}$$

Here Z_L is the load impedance and Z_0 is the characteristic impedance of the transmission line. The line is said to be matched on a reflection basis (or more simply, *reflection-matched*) if $\rho = 0$ (i.e., if $Z_L = Z_0$).

In the most general situation, reflection matching does not ensure maximum power transfer to the load.^{2,3} However, the lines commonly used in antenna systems have very low loss and characteristic impedances that are purely or nearly real. Characteristic impedances designated by Z_0 in this chapter will be assumed to be real unless otherwise noted. In this case, maximum power transfer will occur when the circuit is adjusted for a conjugate-impedance match (Figure 52-1c).

Losses in a transmission line will cause attenuation in the form of reduction in amplitude by a factor $e^{-\alpha\ell}$ for each voltage wave after traveling a length ℓ along a section of line (α is called the attenuation constant of the line, measured in units of Np/m). Reduction in power is thus $e^{-2\alpha\ell}$, which is often expressed in terms of the normal line attenuation $A = 8.686\alpha\ell$ (dB). If the line has nonzero attenuation, maximum power transfer from the source to the load is obtained only when the following conditions are met:

- The generator is loaded by the conjugate of its internal impedance.
- The line is terminated in its characteristic impedance.

The first condition provides maximum power delivery from the source. The second condition provides minimum power dissipation in the line by eliminating the reflected wave on the line. Figure 52-1d shows a “matched” system, that is, a system matched both on a conjugate-impedance basis and on a characteristic-impedance basis. This represents an ideal condition.

The importance of matching can be seen by an examination of the detrimental effects of a mismatch. The power in the load is reduced from the maximum available power by the factor $(1 - |\rho|^2)$. This reduction in transmission is called *reflection loss* [in dB, $-10 \log_{10}(1 - |\rho|^2)$]. Sometimes this effect is expressed in terms of the *return loss* $| \rho |$ (in dB, $-20 \log_{10} |\rho|$). The reflected wave combines with the incident wave on the line to form a

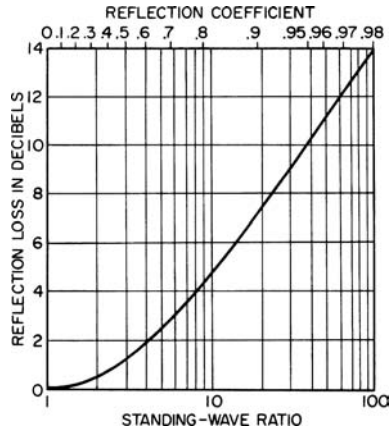


FIGURE 52-2 Reflection loss

standing wave having a standing-wave ratio (SWR) S defined as the ratio of maximum voltage to minimum voltage:^{*}

$$S = \frac{1 + |\rho|}{1 - |\rho|}$$

The presence of the standing wave decreases the breakdown-limited power deliverable to the load by the factor $(S + 1)^2 / (2S)^2$. Transmission loss in dB is shown as a function of the SWR in Figure 52-2.

The efficiency η of the line is the ratio of power delivered to the load to power input to the loaded section of the transmission line. It is given in terms of the SWR and normal line attenuation as

$$\eta = \frac{P_L}{P_{in}} = \frac{4S}{10^{0.1A}(S+1)^2 - 10^{-0.1A}(S-1)^2}$$

where A is the normal line attenuation in decibels.

The Smith Chart

The reflection coefficient is a complex quantity, related directly to impedance ratio and mismatch. Its magnitude does not exceed unity on a lossless line, so a plot of all possible reflection ratios for passive impedances may be charted within a circle of unit radius (see Figure 52-3a). Furthermore, the loci of a mismatch at different distances along the length of a line lie on a constant radius from the center of the chart, at angles proportional to the distances. Coordinate systems of immittance normalized to the characteristic or reference immittance are superimposed to give the useful Smith chart,⁴ shown in skeleton form in Figures 52-3b and c (the characteristic admittance $Y_0 = 1/Z_0$). Additional scales are frequently provided. Interpretation is aided by labels for open- and short-circuit points, and for inductive and capacitive sides.

^{*}The notation S for standing-wave ratio should not be confused with that of a scattering parameter.

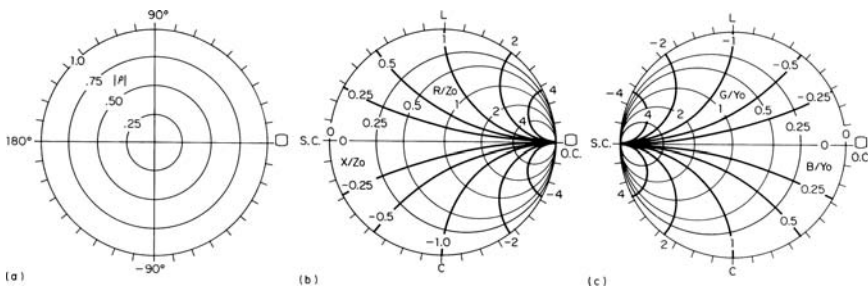


FIGURE 52-3 Reflection coefficient charts in skeleton form, positive real axis to the right: (a) reflection coefficient in polar coordinates $|\rho|$ and ϕ , (b) the Smith chart, with impedance coordinates $r = R/Z_0$ and $x = X/Z_0$, and (c) admittance version of the Smith chart, with coordinates $g = G/Y_0$ and $b = B/Y_0$

Bandwidth Limitations

In principle, frequency-independent resistances can be matched over an arbitrarily wide frequency band. In general, narrowband networks designed to match greatly differing resistances can be replaced by cascades of similar networks, each of which transforms the impedance by a small fraction of the total resistance ratio. Although in practice losses and space considerations will limit how far this approach can be taken, simple networks like the L-sections or quarter-wave transformers, discussed in the following sections, can often be broadbanded in this way.

It is not possible to match a frequency-dependent impedance at all frequencies.⁵ The smaller the desired reflection coefficient, the smaller the bandwidth over which this condition can be realized. While specific matching network topologies are presented in this chapter, it is possible to obtain in a more systematic way networks that have optimum matching behavior over a given range of frequencies. Modern computer-aided design techniques can be employed to advantage in such cases, and the reader is referred to the literature for details.⁶⁻¹⁴

52.2 IMPEDANCE MATCHING WITH LUMPED ELEMENTS

Throughout the lower range of radio frequencies it is convenient to use lumped reactance elements, such as coils and capacitors, in impedance-matching networks. In higher frequency ranges, pure inductance or capacitance may not always be obtained from practical elements, but it is convenient to analyze circuits in terms of their equivalent lumped reactance (and resistance) elements.

Any two complex impedances may be matched by a simple L section of two reactance elements. If, in addition to matching, it is necessary to maintain given phase relationships between the source and load voltages and currents, a third element must be used to form a T or π section. A lattice section using four elements is more convenient for some applications.

The primary use of the simple sections is for matching at a single frequency, although it is possible to obtain matching at two or more separate frequencies by replacing each reactance element of the basic network with a more complex combination, giving the required reactance at each specified frequency. Harmonic reduction can be achieved in a similar manner by introducing high series reactances or low shunt reactances at the harmonic frequencies.

The frequency performance of a matching network is characterized by the fractional bandwidth

$$w = \frac{f_+ - f_-}{f_0}$$

where f_0 is the design frequency, while $f_- < f_0$ and $f_+ > f_0$ are the lowest and highest frequencies at which some specified performance criterion is satisfied.

L Section

The expressions for the required reactance values for an L section providing a match between pure resistances are shown in Figure 52-4a. (Here and in the remainder of this chapter, X and B will designate reactance and susceptance values respectively.) If X_s is positive, the network will delay the phase of the wave by an angle θ . If X_s is negative, it will advance the phase by that angle. The reflection coefficient of R_1 with respect to R_2 is reduced by a factor of at least 10 over a fractional bandwidth of at least 10% by introducing the proper L section of one capacitor and one inductor, provided that $R_2/R_1 < 2.0$.

If two complex terminations are to be matched, the series reactance X_s' is made to include a compensation for the series reactance of the left-hand termination and the shunt susceptance B_p' is made to include the susceptance of the right-hand termination, as shown in Figure 52-4b. For this case θ represents the phase between the current in the left-hand termination and the voltage on the right-hand termination.

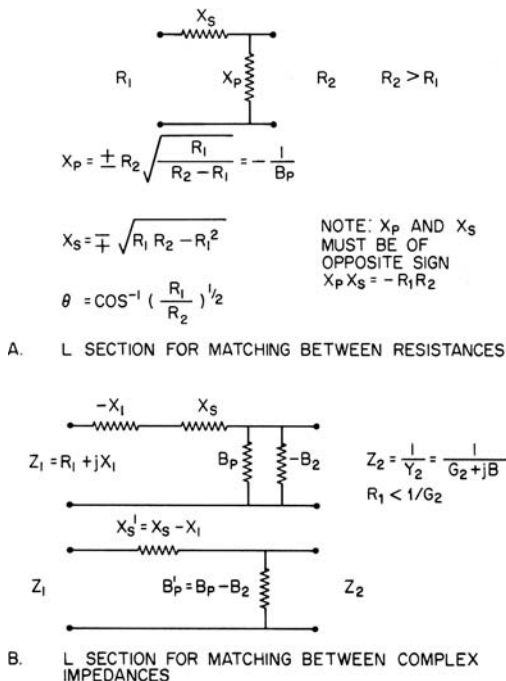


FIGURE 52-4 Impedance matching with reactive L sections

T and π Sections

Design expressions are given in Figure 52-5 for T and π sections matching between resistive terminations with a desired phase shift of θ . If the terminations contain reactive or susceptive components, it is necessary to include appropriate compensating components in the end elements, as in the case of the L section.

L, T, and π sections may be used in unbalanced circuits as shown, or in balanced circuits by moving one-half of each series element to the opposite conductor of the line. The lattice section described next is inherently balanced.

Lattice Section

The elements of a lattice section as well as of T and π sections are given in Figure 52-6 in terms of an equivalent transmission line having the same characteristic impedance and phase shift. The required equivalent line parameters Z_0 and $\theta = \beta l$ may be determined as described in Section 52.3.

Inductive Coupling

A pair of inductively coupled coils is useful in a wide variety of impedance-matching circuits. Figure 52-7 shows two possible equivalent circuits of lossless coupled coils. The first is expressed in terms of reactance elements, including the mutual reactance X_m . The second is in terms of susceptance elements including the transfer susceptance B_T . (The transfer susceptance is the susceptance component of the transfer admittance, which is the ratio of the current induced in the short-circuited secondary to the voltage applied to the primary.) If capacitive tuning reactances are added, the equivalent circuits can be designed in accordance with the T and π matching sections described previously.

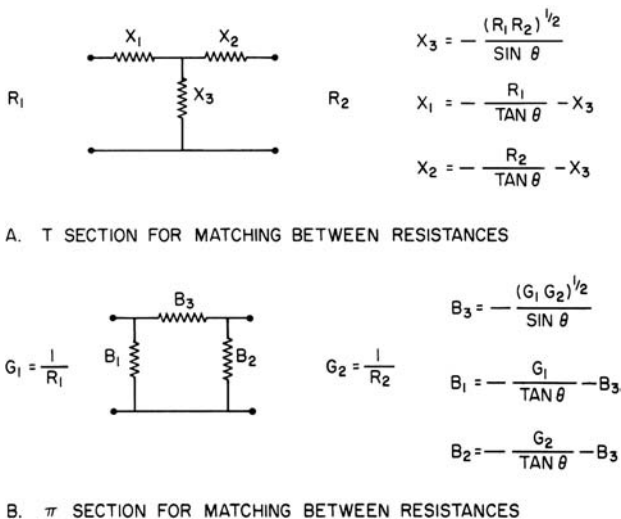


FIGURE 52-5 Impedance matching with reactive T and π sections

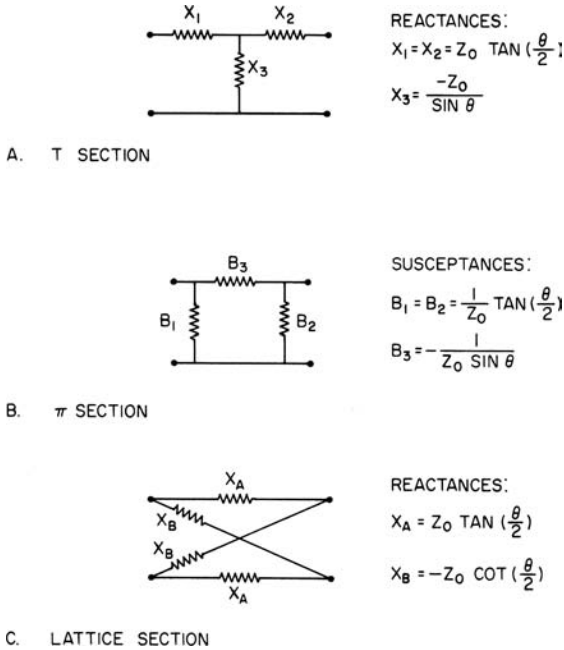


FIGURE 52-6 T, π , and lattice equivalents of a transmission-line section of characteristic impedance Z_0 and electrical length $\theta = \beta\ell$

The capacitive reactances required for (1) series tuning and (2) parallel tuning are given in Figure 52-8 for matching from R_1 to R_2 .

The series and parallel capacitors may also be used for tuning out the series reactance or shunt susceptance, respectively, of a complex termination.

Lumped Matching Reactance

The standing wave on a transmission line may be eliminated on the source side of the load using a matching reactance that is properly chosen and positioned, as shown in Figure 52-9.

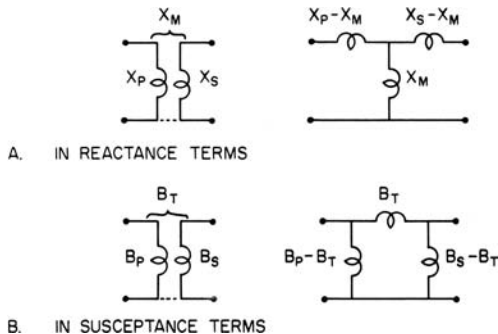


FIGURE 52-7 Inductively coupled circuits and equivalent T and π sections

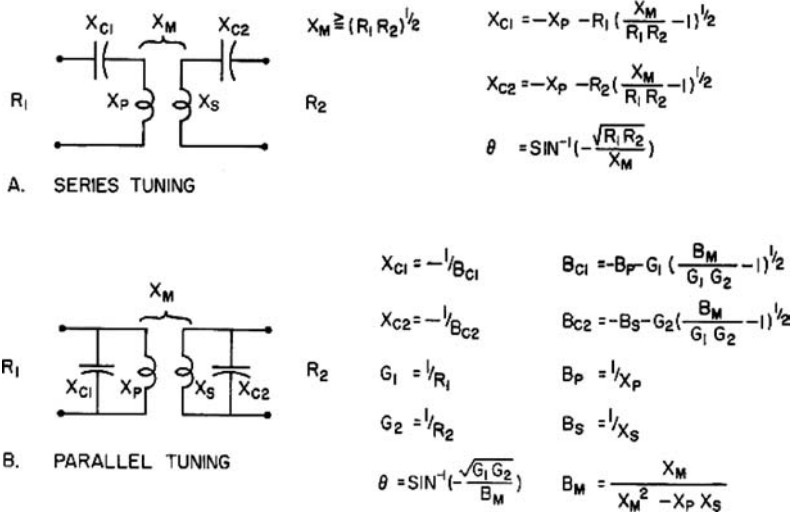


FIGURE 52-8 Tuned inductively coupled circuits for matching between resistances

The choice of series or parallel connections, or inductive or capacitive elements, may be determined by the position of the standing-wave pattern and other practical considerations. A short-circuited or open-circuited transmission-line stub may be used in place of the lumped reactance as described in the next section, resulting in the so-called single-stub matching circuit.

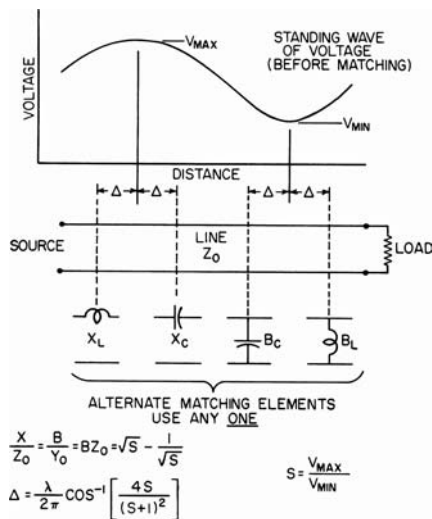


FIGURE 52-9 Determination of lumped reactance to match a transmission line

52.3 IMPEDANCE MATCHING WITH DISTRIBUTED ELEMENTS

In many frequency ranges, it is desirable to use sections of transmission line having distributed reactances rather than lumped reactances in the form of coils and capacitors. The lines usually have negligible loss so that the following expressions for impedance relationships in lossless lines apply:

$$Z_{in} = Z_0 \frac{Z_L \cos \beta \ell + jZ_0 \sin \beta \ell}{Z_0 \cos \beta \ell + jZ_L \sin \beta \ell} \quad (52-1)$$

= input impedance of line ($Z_0, \beta \ell$) terminated in Z_L

For a short-circuited line, $Z_L = 0$, so that

$$Z_{in} = Z_{sc} = jZ_0 \tan \beta \ell \quad (52-2)$$

For an open-circuited line, $Z_L = \infty$, so that

$$Z_{in} = Z_{oc} = -jZ_0 \cot \beta \ell \quad (52-3)$$

It should be kept in mind that when an actual transmission line is connected to a load or to another transmission line, field fringing near the junction point results in excess or *parasitic* inductances and capacitances,¹⁵ whose effects must be compensated for when designing a matching network. Such parasitics will not be explicitly accounted for here, but are readily accounted for in RF circuit modeling software.

Transmission-Line Stubs

Lengths of transmission line short-circuited or open-circuited at one end are often used as reactances in impedance-matching circuits. Inspection of Eqs. 52-2 and 52-3 shows that the choice of characteristic impedance and line length gives the designer control over the reactance value and slope (with respect to frequency) at any given frequency. Alternatively, the designer may control the values of reactance at any two frequencies. If $\beta \ell < \pi/2$, the available slope of reactance is always greater than that obtainable from a single lumped reactance element. Unfortunately, the available slope of reactance is always positive, although a negative slope would be ideal in many applications. However, an effect similar to that of a series-connected or parallel-connected reactance having the unattainable negative slope may often be obtained over a limited frequency range by the use respectively of a parallel-connected or series-connected reactance.

The use of stubs in place of lumped reactances in the matching scheme of Figure 52-9 results in the so-called single-stub matching circuit. Two or three stubs connected at appropriate locations provide additional design flexibility that can be used to increase the bandwidth of the match.¹⁶

Kuroda Identities

In certain cases, the use of either series or shunt-connected stubs may be impractical. In such cases, the Kuroda identity^{17,18} may be employed to replace a shunt-connected open-circuit stub with a series-connected short-circuit stub, at a different location on the line, as shown in Figure 52-10.

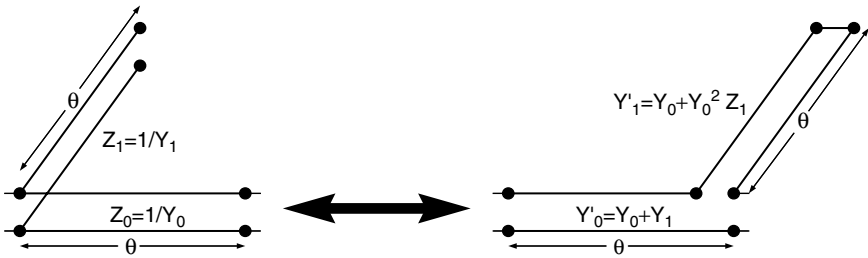


FIGURE 52-10 Kuroda identity

The stubs have the same electrical length, but the characteristic impedances are in general different, as shown.

General Line Transformer

The expression for Z_{in} given in Eq. 52-1 may be recast to relate generator and load impedances ($Z_G = R_G + jX_G$, $Z_L = R_L + jX_L$) to the characteristic impedance and electrical length of a line section, providing a perfect match between them. The new expressions are

$$Z_0 = \sqrt{\frac{R_L |Z_G|^2 - R_G |Z_L|^2}{R_G - R_L}}$$

$$= \sqrt{|Z_G G_L| \frac{\frac{R_L}{R_G} \left| \frac{Z_G}{Z_L} \right| - \left| \frac{Z_L}{Z_G} \right|}{1 - \frac{R_L}{R_G}}}$$
(52-4)

$$\tan \beta \ell = \frac{Z_0 (R_L - R_G)}{R_L X_G - R_G X_L}$$

$$= \frac{Z_0 (X_G - X_L)}{R_L R_G + X_L X_G - Z_0^2}$$
(52-5)

For Z_0 to be a positive finite real number, it is necessary that

$$\frac{1}{2} < \frac{\log \left| \frac{Z_L}{Z_G} \right|}{\log \left(\frac{R_L}{R_G} \right)} < \infty$$

Instead of using the previous expression for $\tan \beta \ell$, one may determine $\beta \ell$ from a plot of Z_L/Z_0 and Z_G^*/Z_0 on a Smith chart once Z_0 has been determined. The two points will be at the same radius. The electrical angle measured clockwise from Z_L/Z_0 to Z_G^*/Z_0 is $\beta \ell$. (Note that $Z_G^* = R_G - jX_G$.) The values of Z_0 and $\beta \ell$ for the required line section may be used to compute T , π , or lattice sections of lumped reactances to perform the same function by using the expressions given in Figure 52-6.

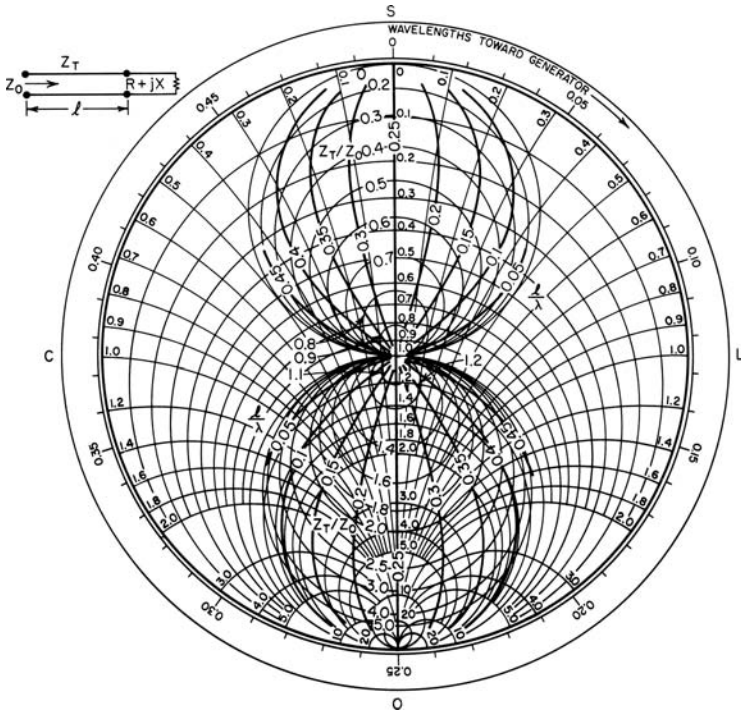


FIGURE 52-11 Line transformer to match to a resistance

Line Transformer for Matching to Resistance

For the frequently encountered case in which either the load impedance or the source impedance is a pure resistance, it may be possible to obtain a solution for the required matching line from Figure 52-11. The complex impedance $R + jX$ is located on the Smith chart in terms of its normalized components R/Z_0 and X/Z_0 , where Z_0 is the resistance to which a match is desired. A second set of coordinates Z_T/Z_0 and ℓ/λ is then overlaid on the Smith chart, corresponding to the characteristic impedance and length of the required matching section of line. If the impedance $R + jX$ is located within the area covered by these new coordinates, a solution is possible. Note that any impedance may be converted to a pure resistance at a single frequency by connecting it to an appropriate length of transmission line. However, if the line length is significant, undesirable frequency dependence may be incurred when using this method.

Quarter-Wave Transformer

The useful quarter-wave transformer results from the general line transformer when $\beta\ell = \pi/2$. It has an impedance-inverting property, as seen from an inspection of

$$Z_{in} = \frac{Z_0^2}{Z_L}$$

The input impedance is thus proportional to the reciprocal of the load impedance. The phase angle of the input impedance is the negative of that of the load impedance. The quarter-wave line can be used, for example, to transform an inductive low impedance to a capacitive high impedance.

The quarter-wave transformer is often used to match between different resistance levels. In this case,

$$Z_0 = \sqrt{R_1 R_2}$$

In a mismatched coaxial line, a quarter-wave transformer formed by a simple external conducting sleeve can be used to tune out the reflection. The sleeve forms an enlargement of the inner conductor or a constriction of the outer conductor, which may be stationed where required. The characteristic impedance of the line is reduced over the quarter-wave length by the sleeve to a value of

$$\frac{Z_0}{\sqrt{S}}$$

where S is the initial SWR of the line. The load end of the sleeve is positioned at a voltage minimum on the standing-wave pattern. The impedance, here looking toward the load, is Z_0/S . The transformer transforms this to Z_0 so that the line on the source side is perfectly matched.

Z_0 Trimmer

The transformation ratio of a quarter-wave section can be adjusted by a reactive trimming immittance, as shown in Figure 52-12, in order to obtain a match when other adjustment of circuit parameters values is not convenient. The immittance required is given by

$$\frac{X}{Z_0} = \frac{R_1 + R_2}{\sqrt{R_1 R_2}} \left(\frac{R_1 R_2}{Z_0^2} - 1 \right) \quad \text{or} \quad \frac{B}{Y_0} = \frac{R_1 + R_2}{\sqrt{R_1 R_2}} \left(\frac{Z_0^2}{R_1 R_2} - 1 \right) \quad (52-6)$$

and the positioning parameter Δ is given by

$$\Delta = \frac{A\lambda}{4\pi} \quad \text{where} \quad A = \cos^{-1} \left(\frac{R_2 - R_1}{R_2 + R_1} \right) \quad (52-7)$$

The resulting network is nearly equivalent to a line for which $Z_0^2 = R_1 R_2$.

Frequency Sensitivity of Line Transformer

The mismatching effect of a departure from the design frequency for the general line transformer (and hence the quarter-wave transformer) can be estimated from Figure 52-13.

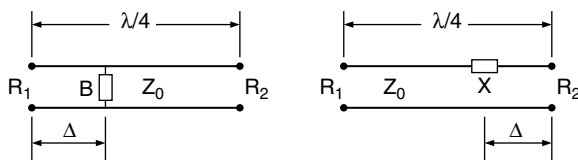


FIGURE 52-12 Shunt and series reactive transformation trimmers for a quarter-wave transformer

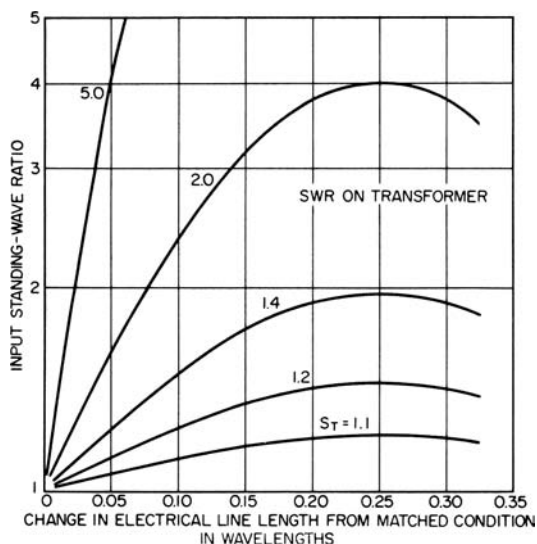


FIGURE 52-13 Frequency sensitivity of a transmission-line transformer

Cascaded Quarter-Wave Transformers^{16,19-21}

A number of quarter-wave transmission-line sections may be arrayed in cascade, in order to realize great improvement in wideband performance over a single-section transformer, as shown in Figure 52-14. The characteristic impedances of the successive sections are proportioned to divide the overall transformation systematically.

In the discussion that follows, $R_1 < R_2$ are the terminating resistances, $R_0 = \sqrt{R_1 R_2}$ is the mean impedance, $f_0 = (f_+ + f_-)/2$ is the design frequency, f_+ is the upper frequency limit, f_- is the lower frequency limit, the frequency ratio is $f_+ / f_- = (2 + w) / (2 - w)$, and w is the bandwidth. Each section is a quarter wavelength long at f_0 .

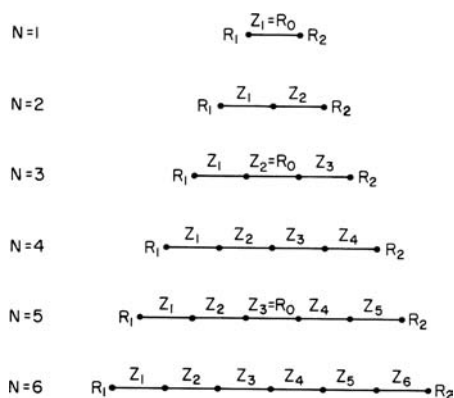


FIGURE 52-14 Cascaded quarter-wave transformers

TABLE 52-1 Design Ratios for Binomial Transformers

$N = n =$	1	2	3	4	5	6	7
1	$1/2$						
2	$1/4$	$3/4$					
3	$1/8$	$4/8$	$7/8$				
4	$1/16$	$5/16$	$11/16$	$15/16$			
5	$1/32$	$6/32$	$16/32$	$26/32$	$31/32$		
6	$1/64$	$7/64$	$22/64$	$42/64$	$57/64$	$63/64$	
7	$1/128$	$8/128$	$29/128$	$64/128$	$99/128$	$120/128$	$127/128$

$$\text{Ratios} = \frac{\log(Z_n/R_1)}{\log(R_2/R_1)} = 1 - \frac{\log(R_2/Z_n)}{\log(R_2/R_1)}$$

Binomial Transformer

The binomial (or binomial-coefficient) distribution gives almost maximally-flat performance. In this distribution the logarithms of the impedance ratios of the steps between sections are made to be in the ratio of the binomial coefficients. Table 52-1 may be used to determine the characteristic impedance Z_n of the n th section in an N -section binomial transformer as a function of R_2/R_1 for values of N up to 7.

The input SWR of an N -section binomial transformer is given by

$$S = 1 + (\cos \theta)^N \ln \frac{R_2}{R_1}$$

where θ is the electrical length of each section at the operating frequency. This expression is subject to the assumptions of small impedance steps, no parasitics at the steps, and equal lengths of the sections.

The two-section transformer has a maximally-flat bandwidth curve for all transformation ratios. For other values of N the performance approximates the maximally flat curve for transformation ratios near unity.

Chebyshev Transformer

If a certain maximum reflection coefficient ρ_m may be tolerated within the operating band, an optimum design in which the reflection coefficient oscillates between 0 and ρ_m within the band and increases sharply outside the band is possible. This is called the Chebyshev transformer, since Chebyshev polynomials are used in its design.

The input SWR of the Chebyshev transformer is given by

$$S = 1 + \ln \left(\frac{R_2}{R_1} \right)^{\frac{T_N \left(\frac{\cos \theta}{\cos \theta_-} \right)}{T_N \left(\frac{1}{\cos \theta_-} \right)}}$$

where $T_m(x)$ is the Chebyshev polynomial of m th degree defined by

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_2(x) = 2x^2 - 1$$

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x)$$

and where θ is the electrical length of each section and θ_- is the electrical length at f_- . The expression is subject to the same assumptions that were made in the binomial case.

Another expression of Chebyshev transformer performance, which does not require the assumption of small steps, is

$$P_L = 1 + \frac{\left(\frac{R_2}{R_1} - 1\right)^2}{\frac{4R_2}{R_1}} \frac{T_N^2\left(\frac{\cos \theta}{\cos \theta_-}\right)}{T_N^2\left(\frac{1}{\cos \theta_-}\right)}$$

where P_L is the power-loss ratio defined by $1/(1 - |\rho|^2)$.

The previous equation may be converted to an implicit relationship among four principal quantities, namely:

R_2/R_1 the transformation ratio,

ρ_m , the maximum tolerable reflection coefficient within the band,

N , the number of sections, and

$F = w/2$, the half-bandwidth or frequency coefficient.

Any one of these may be determined by reference to Figure 52-15, if the other three are known. The design constant C of Figure 52-15 can be used as a measure of the difficulty of transformation. It may be expressed as a function of R_2/R_1 and ρ_m or as a function of N and F :

$$C(R_2/R_1, \rho_m) = (R_2/R_1 - 1) \sqrt{(1/\rho_m^2 - 1) R_1/4R_2}$$

$$C(N, F) = T_N(1/\cos(\pi/2(1 - F))) \geq C(R_2/R_1, \rho_m)$$

The characteristic impedances may be calculated from functions given in Table 52-2. The impedance of the center section when N is odd is simply $\sqrt{R_1 R_2}$ and is independent of F .

Dual-Band Transformers²²

Multisection transformers can be synthesized for dual or multiple passbands by the systematic distribution of the available reflection-coefficient zeros over the passbands. The sequence of impedance levels used in each section becomes nonmonotonic for a large frequency ratio between the two passbands. A nonmonotonic design for one operating band can be made with an arbitrarily short length by placing a second dummy passband at an arbitrarily high frequency.

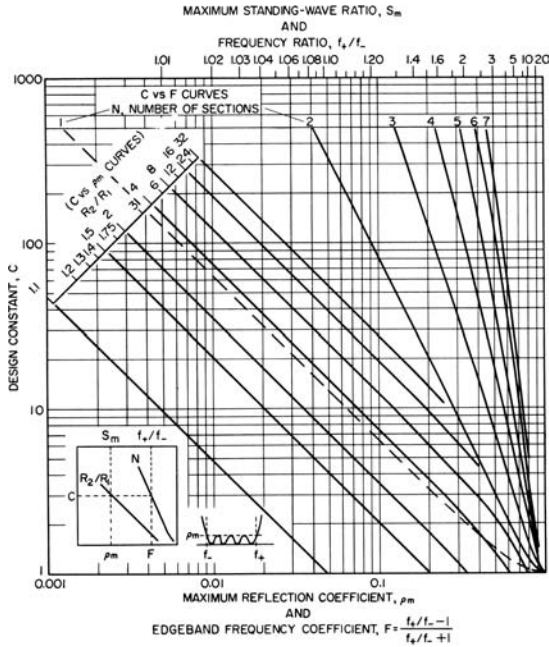


FIGURE 52-15 Performance of a Chebyshev transformer

Short-Step Transformer

A two-section transformer, as shown for $N = 2$ in Figure 52-14, but with a much shorter overall length at midband, is useful for narrow bandwidth or for transformation ratios near unity. $Z_1 Z_2$ is set equal to $R_1 R_2$. The electrical length of each section at the exact-match or design frequency f_0 is set to

$$\beta \ell = \tan^{-1} \sqrt{\frac{R_2/R_1 - 1}{(Z_1/R_1)^2 - Z_2 - Z_1}} \quad (52-8)$$

TABLE 52-2 Design Ratios for a Chebyshev Transformer

N	$\frac{\log (R_2/R_1)}{\log (Z_1/R_1)}$	$\frac{\log (Z_2/R_1)}{\log (Z_1/R_1)}$	$\frac{\log (Z_3/R_1)}{\log (Z_2/R_1)}$
2	$4 - 2C_1^2$		
3	$8 - 6C_1^2$		
4	$16 - 16C_1^2 + 2C_1^4$	$5 - 4C_1^2$	
5	$32 - 40C_1^2 + 10C_1^4$	$6 - 5C_1^2$	
6	$64 - 96C_1^2 + 36C_1^4 - 2C_1^6$	$7 - 6C_1^2$	$32 - 30C_1^2 + 9C_1^4$

NOTE: $C_1 = \cos \theta = \cos [(1 - F) \pi/2]$

The length can be made as short as desired by selecting Z_1/Z_2 sufficiently high within limits imposed by increasing effects of parasitics. Within the ranges $R_2/R_1 < 2$ and $Z_1/Z_2 < 2$, the total bandwidth $2\Delta f$ is equal to $k f_0 \rho_m / \rho_0$, where ρ_m / ρ_0 is the ratio of maximum inband reflection coefficient to the reflection coefficient without the transformer, and k is a coefficient greater than unity. The coefficient becomes much greater than unity for $\beta \ell = \pi/8$ or more. The special case with $Z_1 = R_2$ and $Z_2 = R_1$ is a handy design for matching two dissimilar transmission lines simply by the use of sections of those lines.^{23,24} For this special case, $\beta \ell \leq \pi/6$ and $k \geq 1.1$.

Much advanced theory has been developed for lumped-constant circuits. Useful adaptation to distributed-constant circuits is often possible by transformation of the frequency variable.²⁵

52.4 COMBINATIONS OF TRANSFORMERS AND STUBS

The simple quarter-wave transformer has rather poor performance over a wide band, especially if a large impedance ratio must be matched. It is possible to enhance its performance by combining it with suitably connected stubs as detailed in this section.

Transformer with Two Compensating Stubs

The wideband performance of a quarter-wave transformer may be improved greatly by the addition of a compensating stub at each end, as shown in Figure 52-16. The reactance introduced by each stub counteracts the variation with a frequency of one-half of the transformer length at and near the design frequency. For the impedance proportions shown, the performance is maximally flat and comparable with that of a three-section binomial transformer.

Equiripple performance within the band, for a maximum reflection coefficient of ρ_m , is obtained with modified stub impedances. Table 52-3 lists some examples of equiripple performance, showing an improvement in bandwidth by a factor of about 1.2 over the maximally flat case for $\rho_m = 0.1 \rho_0$.

Transformer with Single Compensating Stub

If it is convenient to use only one compensating stub with the transformer, one of the circuits shown in Figure 52-17 may be used. In this equiripple design, the allowable maximum standing-wave ratio S_m affects the choice of Z_1 and Z_2 .

TABLE 52-3 Transformer with Two Compensating Stubs, Equiripple Case, $\rho_m = 0.1 \rho_0$

R_2/R_1	$Z_2/Z_3 = Z_3/Z_1$	Design F
2	0.6572	0.515
4	2.1205	0.351
6	3.6308	0.286
8	5.1574	0.248

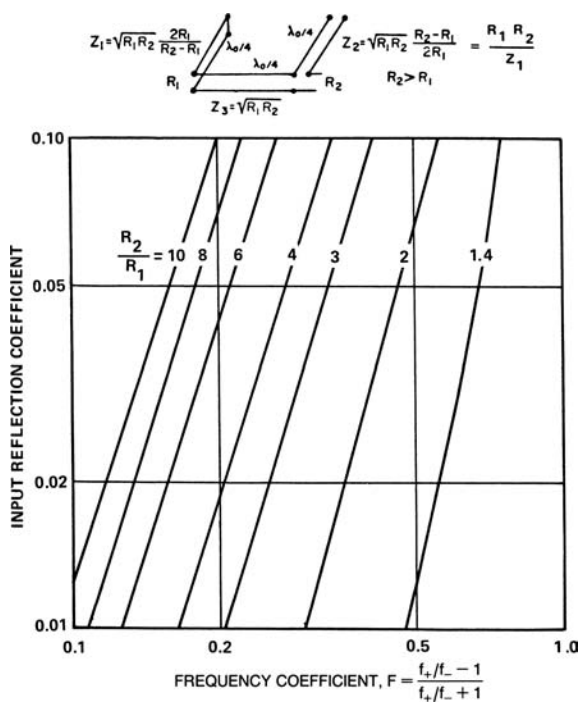


FIGURE 52-16 Transformer with two compensating stubs; maximally flat case

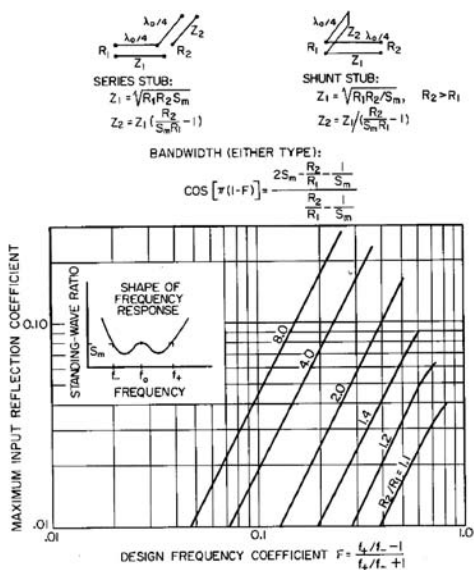


FIGURE 52-17 Transformer with single compensating stub

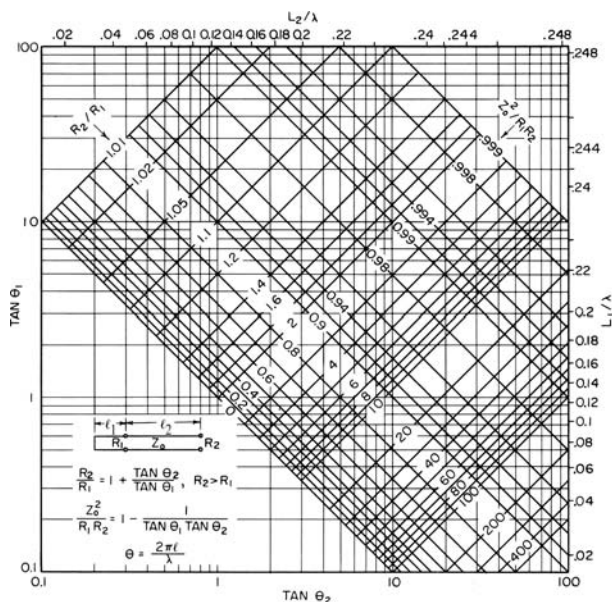


FIGURE 52-18 Tapped-stub transformer

Tapped-Stub Transformer

The tapped-stub transformer, shown in Figure 52-18, is a useful circuit for matching between two widely different resistive terminations. The design chart shows the relationship existing between R_2/R_1 , Z_0^2/R_1R_2 , and the two line lengths ℓ_1 and ℓ_2 for the matched condition. The total stub length is a minimum of one-quarter wavelength and increases with increasing Z_0 . The chart is based on a lossless transmission line. For very high transformation ratios, the effect of even small loss is appreciable and must be accounted for separately. For moderate transformation ratios, a 5 or 10 percent bandwidth over which the maximum reflection coefficient is no more than 0.1 ($R_2 - R_1)/(R_2 + R_1)$) usually can be obtained.

52.5 TAPERED LINES

If the characteristic impedance of a long section of transmission line varies gradually enough with distance along the line, a nearly perfect match between resistive terminations may be obtained. Since a line of such length and such small taper can seldom be afforded, much attention has been directed toward the design of tapered lines giving acceptable transforming performance over short lengths and broad bandwidths.^{16,26-41}

A nonuniform line does not, in general, support the independent propagation of direct and reflected waves as on the uniform line. However, the law of impedance transformation by such a line can be obtained approximately if the taper is gradual enough. For example, if the characteristic impedance $Z_0(x)$ and propagation constant $\beta(x)$ are slowly-varying functions of the position coordinate x along the line ($0 < x < l$), which

is connected between a resistive load $R_L = R_2 = Z_0(\ell)$ and a uniform line in $x < 0$ with characteristic impedance $R_1 = Z_0(0)$, then the reflection coefficient at $x = 0$ is given approximately by:

$$\rho(0^-) = \int_0^\ell e^{-2j\varphi(x)} \frac{Z_0'(x)}{2Z_0(x)} dx \quad (52-9)$$

where $\varphi(x) = \int_0^x \beta(x') dx'$ is the so-called *phase-integral function*. Exact evaluation of the impedance transformation of a tapered line is possible only in a few special cases.

Exponential Line

The exponential line is commonly employed to transform between two resistance levels equal to the values of characteristic impedance at the ends. Sections of exponential line are useful both as simple impedance transformers and as cascaded elements in a more complexly tapered line. The name comes from the function describing the magnitude of the characteristic impedance:

$$Z_0(x) = Z_0(0) \exp(\delta x)$$

where δ is the taper constant and x is the length coordinate that is positive in the direction of increasing impedance. We will assume that the propagation constant β does not vary with x , but this will not be true, for example, in structures such as microstrip. Although exact solutions for the impedance transformation properties of this line (and a few others) are known,³³ it is sufficient for most practical purposes to use the approximation of Eq. 52-9:

$$\rho = 0.5 \ln(R_2/R_1) \frac{\sin(\beta\ell)}{\beta\ell} e^{-j\beta\ell}$$

where $\ln(R_2/R_1) = \delta\ell$. Reflection zeros exist for section lengths of $n\pi$ radians, where n is an integer, and successive maxima of $|\rho|$ between these zeroes decrease in magnitude as $\beta\ell$ increases. The accuracy is acceptable if $\beta\ell > \pi/2$, with the error in phase generally greater than the error in magnitude.

Other Tapered Lines

Several tapered lines^{33,38} are attractive for special simplicity in analysis. Optimum high-pass tapers have also been described.^{16,27} The taper of Klopfenstein^{27,37} is optimum in the sense that for a given taper length the input reflection coefficient, as approximated by Eq. 52-9, has the minimum peak magnitude, or that for a specified peak magnitude the taper has the minimum length. The design is based on the multisection Chebyshev transformer, in which the number of sections is increased without limit so as to raise the upper cutoff frequency indefinitely high. The maxima of $|\rho|$ after the first zero have the same magnitude (equal ripple), and the resulting reflection coefficient in the passband is

$$\rho = \frac{0.5 \ln(R_2/R_1) \cos[(\beta\ell)^2 - A^2] e^{-j\beta\ell}}{\cosh A}$$

where $\beta\ell$ is the electrical length and A is a parameter related to the maximum in-band reflection coefficient by

$$\rho_m = 0.5 \ln (R_2/R_1) / \cosh A$$

At the low-frequency cutoff $\beta\ell = A$, the length of the tapered section is $A/2\pi$ wavelengths. The characteristic impedance taper over the section length ℓ as a function of x , measured from the midpoint of ℓ toward the high-impedance (or R_2) end, is

$$Z_0(x, A) = (R_1/R_2)^{1/2} (R_2/R_1)^{1/2 M(x)}$$

where

$$M(x) = -M(-x) = A^2 \phi(2x/\ell, A) / \cosh A$$

$$\phi(2x/\ell, A) = -\phi(-2x/\ell, A) = \int_0^{2x/\ell} \frac{I_1(A\sqrt{1-y^2})}{A\sqrt{1-y^2}} dy; \quad (2x/\ell \leq 1.0)$$

and I_1 is the modified Bessel function of the first kind and first order. Values of the function ϕ can be computed for different values of $z = 2x/\ell$ and A using the simple routine of Grossberg:³⁶

$$\phi(z, A) = \sum_{k=0}^{\infty} a_k b_k$$

where

$$a_0 = 1; \quad a_k = \frac{A^2}{k(k+1)} a_{k-1}; \quad b_0 = \frac{z}{2}; \quad b_k = \frac{\frac{z}{2}(1-z^2)^k + 2kb_{k-1}}{2k+1}$$

and the coefficients a_k and b_k are computed iteratively, the series being truncated when the desired numerical convergence is reached.

Other criteria for choosing an optimal taper are possible as well.^{29-32,34-41} For example, it is possible to minimize reflection only over a certain band of frequencies, at the expense of higher mismatch outside this band.

52.6 BROADBANDING⁴²

The most general problem in broadbanding is that of synthesizing a circuit to match one arbitrary impedance to another arbitrary impedance over a prescribed frequency range to within a prescribed tolerance. It has already been remarked that in general this is a very difficult design problem. Fortunately, the broadbanding problems found in practice involve impedances that are not completely arbitrary, inasmuch as they are composed of a combination of physically realizable inductances, capacitances, resistances, and transmission line sections. A further simplification of the broadbanding problem often occurs because at least one of the impedances to be matched is usually the characteristic impedance of a transmission line and thus is either constant and resistive or nearly so.

Nevertheless, the impedances to be matched are functions of frequency that usually are too complex for simple analysis. An iterative approach is often taken. The subject impedance,

prescribed by theoretical or experimental data, is examined to perceive a suitable step toward optimum broadbanding. Graphical representations of the function are usually helpful for this. The matching network or a portion of it is added to the circuit. The resulting impedance is found by calculation or experiment. The steps are repeated until the required broadband match is obtained. Modern computational tools and measuring instruments minimize the tedious detail and impart the needed accuracy and so make the iterative approach rapid and efficient.

The problem of matching two resistive impedances has been treated in Sections 52.2 to 52.5. The problem of matching a load of frequency-varying impedance to a constant resistance is so common that the following discussion will be limited to this case.

Several broad rules that apply to broadbanding practice are given, as follows:

1. The difficulty of obtaining a prescribed tolerance of match increases with the required bandwidth. Refer to performance charts of a Chebyshev transformer for an example of this (and of the following rule).
2. The difficulty increases with impedance transformation ratio (expressed as a quantity greater than 1).
3. The difficulty increases with the electrical length of the transmission circuit between the load and the first point of control.
4. Improvement of match throughout one sector of the frequency range will generally be accompanied by an increase in mismatch in other sectors of the frequency range.
5. Any physically realizable passive impedance plotted on the reflection-coefficient plane, using the conventions of Figure 52-3, displays a circular or spiral motion having a clockwise sense of rotation with frequency.

The effect of a length of mismatched transmission line between the load terminals and the first impedance-matching control is to introduce an additional variation with frequency that is seldom favorable over an appreciable bandwidth (see rule 3). For this reason it is advantageous to conduct impedance-matching control at a position close to the load terminals. In fact, if control of the load impedance is available from within the load itself, advantage should be taken of this circumstance to select the most suitable shape and position of the impedance locus on the reflection-coefficient plane.

Two things are required in order to perform a transformation of the given impedance locus on the chart to one that is compactly situated about the *desired* impedance point. First, a method of moving the impedance locus, and second, a method of compensating for both the variation with frequency inherent in the original impedance locus and that introduced by the moving process must be found. Means for moving the impedance locus may include lumped or distributed constant elements forming shunt or series reactances, cascade transformers, or a combination of these. The selection is made so as to introduce as little adverse variation of impedance with frequency as possible.

Compensation of variation of impedance with frequency is usually limited to a band including less than one complete convolution of the original impedance locus if a high degree of compensation is required. Typically, a sector of the original impedance locus, which may be made to appear similar to those shown in Figure 52-19, is selected. Then a matching circuit having a variation of reactance of opposite sense (and of the proper magnitude) is added to yield the tightly knotted transformed locus as shown. Essentially all the reactance variation may be eliminated; however, a small reactance variation and a larger resistance variation remain. In certain instances it may be advantageous to utilize two stages of compensation, as illustrated in Figure 52-20, in order to accomplish a doubly knotted transformed locus.

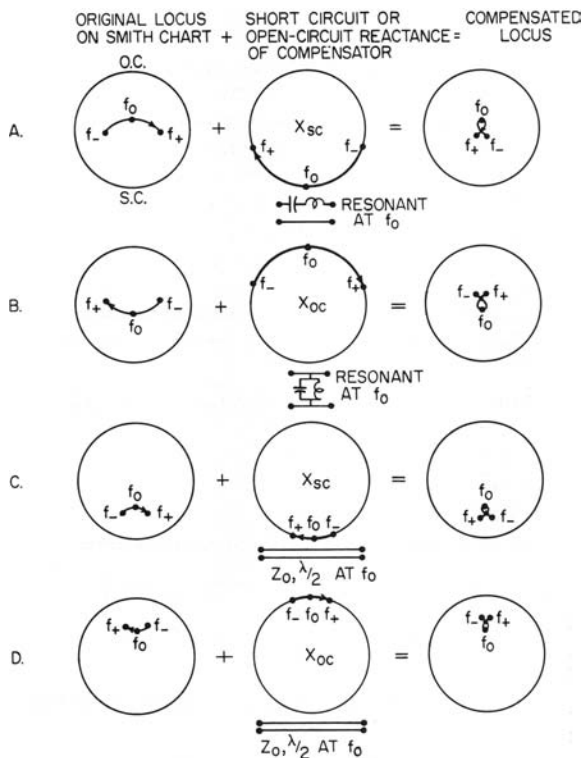


FIGURE 52-19 Broadband compensation methods

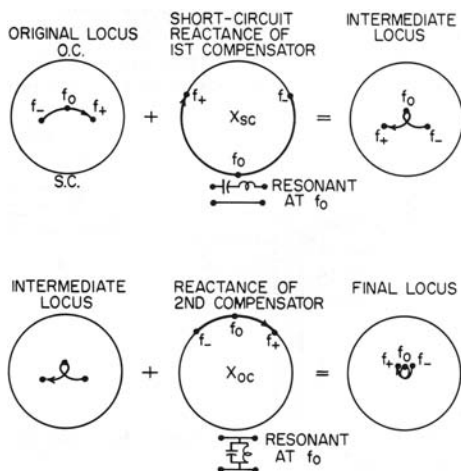


FIGURE 52-20 A two-stage compensation method

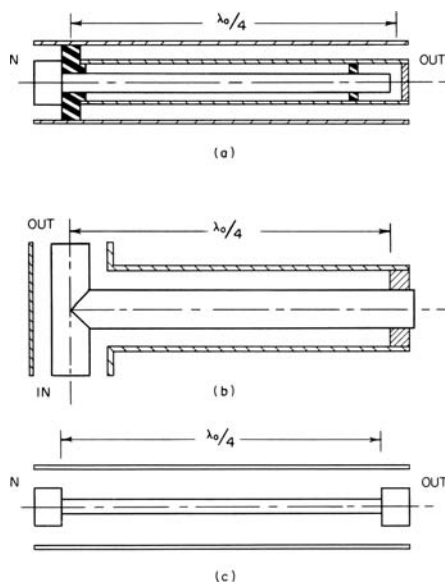


FIGURE 52-21 Possible physical embodiments of Figure 52-19

There are many ways of forming these matching circuits into physical structures. A few simple examples are shown in Figure 52-21. A more complex structure, utilizing a transformer with two compensating stubs in a circuit described earlier in this chapter, is shown in Figure 52-22.

Resonant lengths of the transmission line may be used in place of the lumped-constant resonant circuits. The LC product or the length of the line is selected for resonance at f_0 , the midfrequency of the sector. The L/C ratio or the characteristic impedance of the line section is selected to give the proper amount of compensating reactance at f_+ and f_- , the edge-band frequencies.

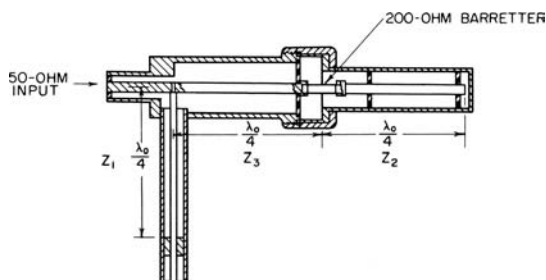


FIGURE 52-22 Structure to match barretter to 50- Ω impedance

52.7 BALUNS

Antennas, if not driven properly, may interact in undesirable and unpredictable ways with their environment. Frequently, this interaction can be minimized by enforcing balanced operation: the currents in the two arms of a symmetric dipole should be equal, for example. A *balun* is a three-port network designed to couple an unbalanced transmission circuit at one port to a balanced transmission circuit at the other two ports. A balun may be a current balun, which forces the currents at the balanced ports to be equal and opposite, or a voltage balun, which does the same thing for the voltages at the balanced ports. Though many types of balun also possess impedance transformation properties, this is not the distinguishing feature of balun operation. Any desired impedance transformation can usually be accomplished using established methods.^{43,44} Conversion between a balanced mode and an unbalanced mode requires special techniques. Many balun designs have been developed, some of the basic types of which are shown in Figure 52-23. Further information on balun design can be found in the references.⁴⁵⁻⁴⁸

A balanced circuit is obtained in type 1 by introducing a high impedance of a resonant coaxial-choke structure between the outer conductor of the unbalanced coaxial circuit and ground. There is no counterpart of the choke impedance in the other side of the balanced circuit. Hence the balun does not present a well-balanced impedance at any frequency for which the choke impedance is not high. This shortcoming of type 1 is overcome in types 2 and 3. Symmetry of the balanced circuit is maintained independently of frequency by a two-conductor choke (type 3) or by two identical, opposed, coaxial-choke cavities (type 2). The first three types have been widely used because of their simplicity and mechanical adaptability. The impedance bandwidth is limited by the shunting effect of the choke. Thus it is advantageous to keep the characteristic impedance of the choke lines as high as possible for a wideband operation.

Type 4 represents an improvement for many applications, particularly those involving impedance-level transformation. Unlike types 2 and 3, type 4 has the balanced circuit in the innermost line. Coupling is attained through a gap between line sections, as before. The necessary choke cavity, however, is around this gap in the unbalanced side of the structure. An unbalanced or single-ended cavity can be used without disturbing the perfect balance. The connections at the gap may be either of the series type, as in types 4a and 4c, or of a parallel type, as in types 4b and 4d. The parallel connection results in a balanced-to-unbalanced impedance ratio of 4.

An additional impedance control is available in any of the type 4 baluns at some expense of bandwidth by application of the principle of the tapped-stub transformer. This is illustrated by the alternative position of the unbalanced line in the drawing of type 4a. The desired impedance transformation ratio determines the length of coaxial-choke line on each side of the unbalanced-line tap position in accordance with Figure 52-18. Broadband compensation techniques⁴³ may be applied to types 2, 3, and 4.

The type 5 balun employs a half-wave delay line. It is connected between the two balanced-line terminals as shown, forcing the potentials to ground to be equal and opposite at the design frequency. A balanced-to-unbalanced impedance ratio of 4 is obtained. Bandwidth can be increased by adding multiple sections.⁴⁹

Type 6 employs helically wound transmission lines. At the unbalanced end, they are connected in parallel. Enough line length is used to develop high impedances to ground at the opposite ends. There the lines are connected in series to form the balanced terminals. A balanced-to-unbalanced impedance ratio of 4 is obtained.

Type 7a is a conventional $6\lambda/4$ hybrid ring which operates similarly to type 5. The path length from A to C is one-half wave longer than the path from A to B. Each path is an odd number of quarter waves long so that impedance transformation may be incorporated by

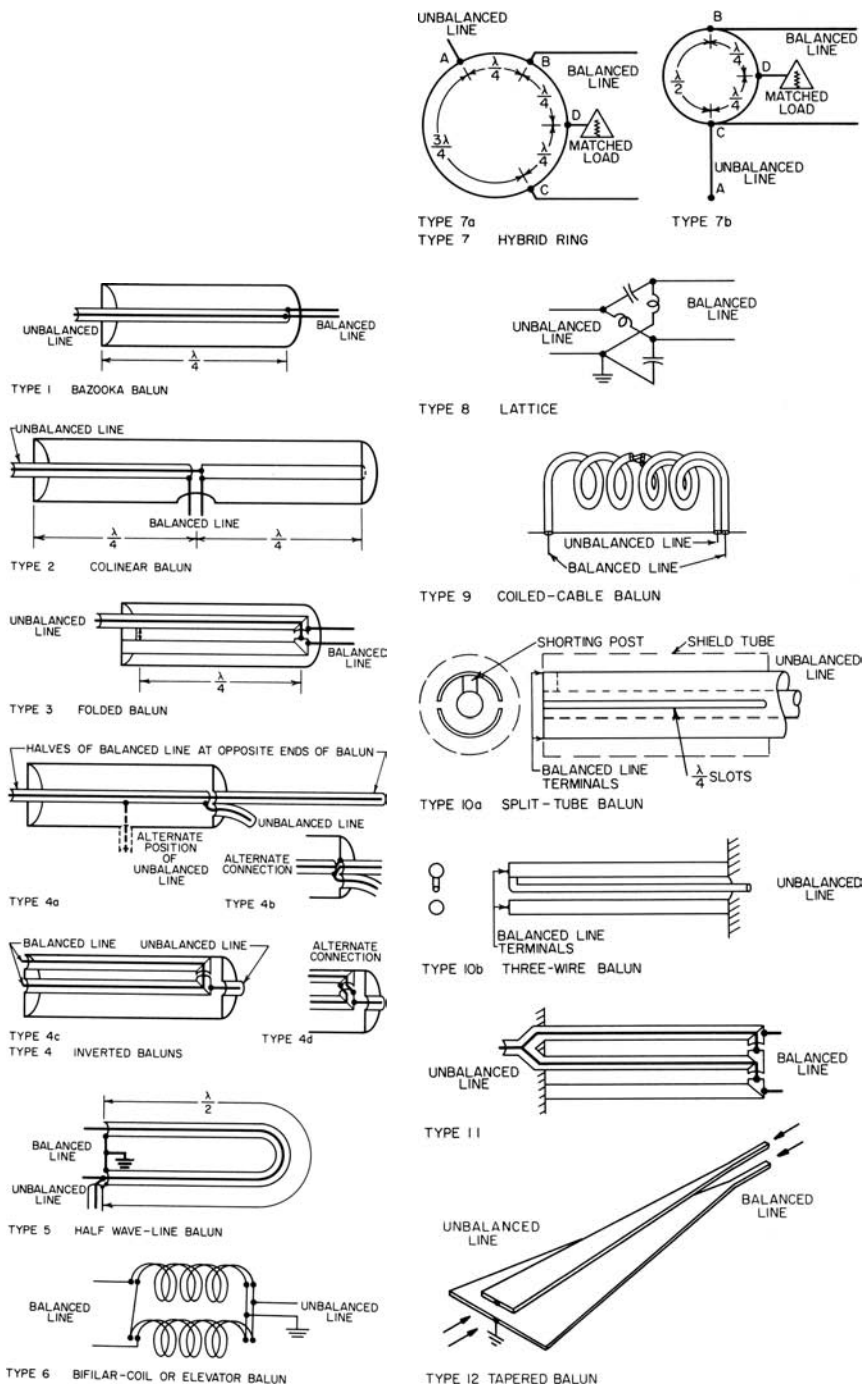


FIGURE 52-23 Types of baluns

proper selection of wave impedances. Furthermore, a load resistor at D is connected to terminals B and C by quarter-wave lines. This arrangement tends to dissipate the energy of an unbalanced wave without affecting the balanced wave. Type 7b is a $4\lambda/4$ hybrid ring that operates similarly, except that the impedance-matching function is performed by the separate quarter-wave line AC .

Type 8 is a familiar lattice circuit, wherein the incoming unbalanced wave energy is divided equally between two channels, one providing a 90° lead, the other a 90° lag. The output voltage is balanced with respect to ground and is in quadrature with the input voltage. The lattice may be proportioned to match any two resistance values. The lumped-element reactances may be replaced with transmission-line stubs, which is particularly suitable for printed-circuit applications.⁵⁰⁻⁵¹ The lattice balun may be generalized to a higher-order circuit to improve its bandwidth.⁵²

Type 9 is similar in principle to type 6. It is most suitable for operation at the longer wavelengths where the size of a type 4 tends to become excessive. It may be connected for an impedance ratio of 1 or of 4.

Type 10a is a convenient balun type for feeding balanced dipoles. The slotted portion of transmission line supports two modes of transmission simultaneously. Energy in a coaxial mode from the unbalanced input is substantially unchanged by the presence of the slots. The field of this mode is almost completely confined within the outer conductor. There is also a balanced mode in which the halves of the slotted cylinder are at opposite potentials and the center conductor is at zero potential. The field is not confined to the space within the slotted cylinder. (If leakage and radiation are to be minimized, another cylinder may be added overall for shielding.) The connection strap between the inner conductor and one-half of the slotted cylinder requires that the coaxial-mode voltage be equal and opposite to that half of the balanced-mode voltage. Thus a balanced-to-unbalanced impedance ratio of 4 is obtained.

Type 10a will give almost perfect balance over a wide frequency range if the slot width is kept small and symmetry is maintained at the strap end. A variation of this basic type of balun is the three-wire balun shown in type 10b. It is sufficiently well balanced for use in noncritical applications.

Type 11 illustrates another method by which an impedance transformation may be obtained by a connection independently of frequency. Two coaxial lines are connected in parallel at the unbalanced terminal and in series at the balanced terminal. A third cylinder is added to preserve symmetry. The impedance transformation ratio for this structure having two lines is 2^2 , or 4. More coaxial lines could be added to give impedance transformation ratios of 9, 16, etc., but practical limits are soon reached in this direction. Bifilar coils can also be used in a balun operating on this principle.

Baluns of types 1-4, 10a,b, and 11 are based on various multiconductor transmission lines. Many variants on this idea have been studied in detail.⁵³⁻⁵⁵ Some types of balun, for example 1 through 4, 6, 9, and 11, can be improved for very wideband operation by the inclusion of such high-permeability materials as ferrite or powdered iron. The material in the form of toroidal cores, cylindrical cores, pot cores, or beads, placed in the magnetic field of a low-frequency-limiting shunt element, increases the length of that element, raises its impedance, and improves the low-frequency match. The influence of the material is much less at the high-frequency end, which can be 1 GHz or higher, because of a reduced effective permeability. Compact baluns and transmission-line transformers of the bifilar-coil and coiled-cable types, having ferrite cores, typically operate over bandwidths of two to three decades.⁵⁶⁻⁵⁷ Core loss is low because the core does not interact strongly with the transmission-line mode.

The wideband balun of type 12 is a half wave or more of tapered transmission line that converts gradually, in cross-sectional characteristics, from an unbalanced line (coaxial line or stripline) at a first impedance at one end to a balanced two-conductor line at a second impedance at the other end, using a Chebyshev taper.⁵⁸⁻⁵⁹

REFERENCES

1. W. L. Everitt and G. E. Anner, *Communication Engineering*, 3rd Ed. (New York: McGraw-Hill, 1956).
2. R. A. Chipman, "Dilemmas in Transmission-line Theory," *Electronic and Radio Engineer*, vol. 34 (February 1957): 64–67.
3. G. Raisbeck, "Optimal Lumped Loading with Four-terminal Networks," *Electronic Technology*, vol. 38 (December 1961): 437–440.
4. P. H. Smith, *Electronic Applications of the Smith Chart*, 2nd Ed. (Atlanta: Noble Publishing, 1995).
5. R. M. Fano, "Theoretical Limitations on the Broadband Matching of Arbitrary Impedances," *J. Franklin Inst.*, vol. 249 (January and February 1950): 57–83 and 139–154.
6. H. J. Carlin, "A New Approach to Gain-bandwidth Problems," *IEEE Trans. Circ. Syst.*, vol. 24 (April 1977): 170–175.
7. H. J. Carlin and P. Amstutz, "On Optimum Broadband Matching," *IEEE Trans. Circ. Syst.*, vol. 28 (May 1981): 401–405.
8. B. S. Yarman and H. J. Carlin, "A Simplified 'Real Frequency' Technique Applied to Broadband Multistage Microwave Amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 30 (December 1982): 2216–2222.
9. H. J. Carlin and B. S. Yarman, "The Double Matching Problem: Analytic and Real-frequency Solutions," *IEEE Trans. Circ. Syst.*, vol. 30 (January 1983): 15–28.
10. B. S. Yarman, "Modern Approaches to Broadband Matching Problems," *IEE Proc.*, part H, vol. 132 (April 1985): 87–92.
11. W.-K. Chen, *Broadband Matching: Theory and Implementations*, 2nd Ed. (Singapore: World Scientific, 1988).
12. S. Rosleniec, *Algorithms for Computer-Aided Design of Linear Microwave Circuits* (Boston: Artech House, 1990).
13. S. E. Sussman-Fort, "The Computer-aided Design of Microwave Matching Networks," *Int. J. Microwave and Millimeter-Wave Computer-Aided Design.*, vol. 1 (July 1991): 288–305.
14. M. W. Medley, Jr., *Microwave and RF Circuits: Analysis, Synthesis and Design* (Boston: Artech House, 1993).
15. K. C. Gupta, R. Garg, and R. Chadha, *Computer-Aided Design of Microwave Circuits* (Dedham, MA: Artech House, 1981).
16. R. E. Collin, *Foundations for Microwave Engineering*, Chap. 5 (New York: IEEE Press, 2001).
17. J. Helszajn, *Synthesis of Lumped Element, Distributed, and Planar Filters*. (London: McGraw-Hill, 1990).
18. H. J. Carlin and P. P. Civalieri, *Wideband Circuit Design* (Boca Raton, FL: CRC Press, 1998).
19. S. B. Cohn, "Optimum Design of Stepped Transmission-line Transformers," *IRE Trans. Microwave Theory Tech.*, vol. 3, no. 3 (April 1955): 16–21.
20. H. J. Riblet, "General Synthesis of Quarter-wave Impedance Transformers," *IRE Trans. Microwave Theory Tech.*, vol. 5 (January 1957): 36–43.
21. G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures* (Dedham, MA: Artech House, 1980).
22. A. G. Kurashov, "Multiband Stepped Taper," *Radio Eng. Electron. Phys.*, vol. 24, no. 2 (February 1979): 10–15.
23. A. G. Kurashov, "A Step Transition Without Intermediate Characteristic Impedances," *Telecommun. Radio Eng.*, vol. 39/40, no. 3 (March 1985): 33–34.
24. V. P. Meshchanov, S. A. Saliy, I. A. Tsots, and L. V. Shikova, "Synthesis of Stepped Transition Units of Class II," *Radio Eng. Electron. Phys.*, vol. 29, no. 12 (December 1984): 9–18.
25. A. I. Grayzel, "A Synthesis Procedure for Transmission Line Networks," *IRE Trans. Circ. Theory*, vol. 5 (September 1958): 172–181.

26. H. Kaufman, "Bibliography of Nonuniform Transmission Lines," *IRE Trans. Antennas Propagat.*, vol. 3 (October 1955): 218–220.
27. R. W. Klopfenstein, "A Transmission Line Taper of Improved Design," *IRE Proc.*, vol. 44 (January 1956): 31–35.
28. E. F. Bolinder, "Fourier Transforms and Tapered Transmission Lines," *IRE Proc.*, vol. 44 (April 1956): 557.
29. J. Willis and N. K. Sinha, "Non-uniform Transmission Lines as Impedance Transformers," *Proc. IEE (London)*, pt. B, vol. 103 (March 1956): 166–172.
30. A. L. Fel'dshtein, "The Design of an Optimum Smooth Transition," *Radio Engineering*, vol. 14, no. 3 (March 1959): 52–61.
31. L. Solymar, "A Note on the Optimum Design of Non-uniform Transmission Lines," *Proc. IEE (London)*, pt. C, vol. 107 (March 1960): 100–104.
32. P. G. Cath, "The Synthesis of Nonuniform Transmission Lines," Tech. Rept. No. 116, Cooley Electronics Laboratory, Department of Electrical Engineering, University of Michigan, January, 1961.
33. R. N. Ghose, *Microwave Circuit Theory and Analysis*, Chap. 12 (New York: McGraw-Hill, 1963).
34. P. I. Kuznetsov and R. L. Stratonovich, *The Propagation of Electromagnetic Waves in Multiconductor Transmission Lines* (New York: Macmillan, 1964).
35. A. H. Hall, "Impedance Matching by Tapered or Stepped Transmission Lines," *Microwave J.*, vol. 9, no. 3 (March 1966): 109–114.
36. M. A. Grossberg, "Extremely Rapid Computation of the Klopfenstein Impedance Taper," *IEEE Proc.*, vol. 56 (September 1988): 1629–1630.
37. D. Kajfez and J. O. Prewitt, "Correction to 'A transmission line taper of improved design'," *IEEE Trans. Microwave Theory Tech.*, vol. 21 (May 1973): 364.
38. M. J. Ahmed, "Impedance and Transformation Equations for Exponential, Cosine-squared, and Parabolic Tapered Transmission Lines," *IEEE Trans. Microwave Theory Tech.*, vol. 29 (January 1981): 67–68.
39. Y. Wang, "New Method for Tapered Transmission Line Design," *Electron. Lett.*, vol. 27 (December 5, 1991): 2396–2398.
40. H. Oraizi, "Design of Impedance Transformers by the Method of Least Squares," *IEEE Trans. Microwave Theory Tech.*, vol. 44 (March 1996): 389–399.
41. E. J. Park, "A General Perturbational Approach for Synthesizing the Riccati-type Nonuniform Transmission Lines with Arbitrarily Specified Reflection Properties," *Micr. Opt. Technol. Lett.*, vol. 38 (July 5, 2003): 52–56.
42. V. H. Rumsey, "The Design of Frequency-compensating Matching Sections," *IRE Proc.*, vol. 38 (October 1950): 1191–1196.
43. G. Oltman, "The Compensated Balun," *IEEE Trans. Microwave Theory Tech.*, vol. 14 (March 1966): 112–119.
44. H. R. Phelan, "A Wide-band Parallel-connected Balun," *IEEE Trans. Microwave Theory Tech.*, vol. 18 (May 1970): 259–263.
45. W. L. Weeks, *Antenna Engineering* (New York: McGraw-Hill, 1968).
46. H. Meinke and F. W. Gundlach, *Taschenbuch der Hochfrequenztechnik*, 3rd Ed. (Berlin: Springer-Verlag, 1968): 389–400.
47. Hu Shuhao, "The Balun Family," *Microwave J.*, vol. 30, no. 9 (September 1987): 227–229.
48. J. Seveck, *Transmission Line Transformers* (Newington, CT: American Radio Relay League, 1990).
49. M. Al Basraoui and P. N. Shastri, "Wideband Planar Log-periodic Balun," *Int. J. RF Microwave Computer-Aided Eng.*, vol. 11 (November 2001): 343–353.
50. V. Trifunović and B. Jokanović, "Four Decade Bandwidth Uniplanar Balun," *Electron. Lett.*, vol. 28 (March 12, 1992): 534–535.

51. V. Trifunović and B. Jakanović, "Review of Printed Marchand and Double Y Baluns: Characteristics and Application," *IEEE Trans. Microwave Theory Tech.*, vol. 42 (August 1994): 1454–1462.
52. D. Kuylenstierna and P. Linnér, "Design of Broad-band Lumped-element Baluns with Inherent Impedance Transformation" *IEEE Trans. Microwave Theory Tech.*, vol. 52, (December 2004): 2739–2745.
53. N. Nagai and A. Matsumoto, "Application of Distributed-constant Network Theory to Balun Transformers," *Electron. Commun. Japan*, vol. 50, no. 5 (May 1967): 114–121.
54. N. Nagai, T. Nishide, and A. Matsumoto, "Multiwire-line Baluns and Unbaluns," *Bull. Res. Inst. Appl. Electricity*, vol. 22, no. 1-2 (1970): 24–51.
55. T. Nishide, N. Nagai, and A. Matsumoto, "Balance-to-unbalance Transformers," *Recent Developments in Microwave Multiport Networks*, A. Matsumoto (ed.) (Sapporo, Japan: Research Institute of Applied Electricity, 1970): 69–85.
56. C. L. Ruthroff, "Some Broad-band Transformers," *IRE Proc.*, vol. 47 (August 1959): 1337–1342.
57. J. Seveck, "Broadband Matching Transformers Can Handle Many Kilowatts," *Electronics*, vol. 49, no. 24 (November 25, 1976): 123–128.
58. J. W. Duncan and V. P. Minerva, "100:1 Bandwidth Balun Transformer," *IRE Proc.*, vol. 48 (February 1960): 156–164.
59. M. Gans, D. Kajfez, and V. H. Rumsey, "Frequency Independent Balun," *IEEE Proc.*, vol. 53 (June 1965): 647–648.

Chapter 53

Radomes

David G. Burks

Raytheon Company

CONTENTS

53.1	GENERAL DISCUSSION	53-2
53.2	RADOME WALL DESIGN	53-5
53.3	ELECTRICAL ANALYSIS AND MODELING APPROACHES	53-12
53.4	RADOME MATERIALS	53-16
53.5	FABRICATION AND TESTING	53-18

53.1 GENERAL DISCUSSION

This section introduces the need for radomes and their purpose in antenna systems. Design considerations for electrical, mechanical, and environmental performance are then discussed.

Purpose of Radomes

A *radome* is a protective enclosure for an antenna.¹ The word radome was coined as a contraction of “radar dome,” however many radomes do not cover radar antennas and are not shaped like domes. The primary function of a radome is to protect an antenna system from the environments encountered in ground-based, shipboard, airborne, and aerospace applications. It is also important that a radome be designed to have minimum impact on the electrical performance of enclosed antennas.

Radomes are generally composed of low-loss dielectrics shells, with thickness comparable to the wavelength of operation. They are shaped to cover the antenna and provide the internal volume necessary for mechanical antenna scanning. If needed for the application, they provide aerodynamic streamlining. Radomes are used with virtually all aperture-type airborne antennas and with many ground-based and shipboard aperture antennas, which must be protected from severe weather conditions. See Figure 53-1.

The importance of protecting antennas in their operational environment was recognized from the early days of radar. Many significant interactions with scanning antennas were then identified.² Stationary and conformal antennas on vehicles also frequently require protection by flat or slightly curved windows. Applications include communication and radar altimeter antennas for reentry spacecraft, where extreme temperatures are encountered and ceramic materials must be used.³ A *radant* is a radome that includes an antenna as an integral part. Today, the term *electromagnetic window* is used for all types of radomes, windows, and radants, as well as for the transparent protective covers for optical and infrared sensors.

Electrical Considerations

The electrical performance of an enclosed antenna is always altered by a radome. This is due to distortion of the fields near the antenna caused by interactions at the material interfaces, and amplitude, phase, and polarization changes in the radome material. Radome-induced pattern distortion (Figure 53-2) includes changes in gain, sidelobe levels, beamwidth, null depth, and polarization characteristics. Streamlined radomes cause pointing or *boresight error* in angle-tracking systems. For proportional navigation systems, the rate of change of boresight error with antenna scan (*boresight-error slope* or rate) will affect guidance accuracy. Reflected power from the radome may cause magnetron pulling.² For high-power applications, excessive losses in the radome material can raise the temperature to a point where both structural properties and electrical performance are degraded.⁴ Radome losses also raise the system noise temperature.⁵

First-order radome effects are often explained in terms of transverse electromagnetic (TEM) wave transmission and reflection from planar dielectric panels. Waves transmitted by large aperture antennas can be approximated as a plane wave near the antenna and a curved radome wall can be considered as locally planar at each point of incidence. Reflected and transmitted waves are then approximated from planar slab theory and the



(a) Aircraft, missile, and infrared electromagnetic windows



(b) Radomes for sea-based radar and communication antennas

FIGURE 53-1 Radomes for airborne and surface based applications (*Courtesy Raytheon Company*)

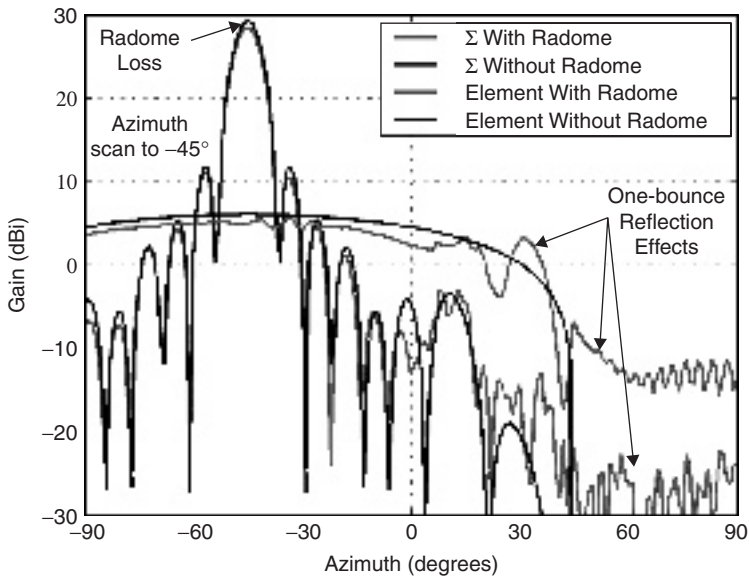


FIGURE 53-2 Radiation patterns showing radome effects on element and sum beams

resultant effects on overall antenna performance calculated. By reciprocity, the effect of the radome on a receiving antenna is the same as for transmitting.

Mechanical Considerations

Radome design is based on the materials and construction needed to provide adequate safety margins and acceptable weight. The materials must provide necessary strength over the service temperature and not have fatigue problems from vibration or impacts.

Mechanical stresses are produced in radomes by aerodynamic loading due to air-flow, acceleration forces, and sudden thermal expansion due to aerodynamic heating (thermal shock).⁶ In high-speed radomes, thermal shock often causes the highest mechanical stresses. These stresses are greatest in the nose region, which often contains a metal tip that must be integrated with the mechanical design.

The radome attachment to the airframe is also a critical design problem in high-temperature applications, because high aerodynamic loads produce large bending moments, which often occur near the end of missile flight and after exposure to significant heating. Bonding, fasteners, and clamps have been used for attachments.

Environmental Considerations

The radome environment is a primary factor in determining material type, wall design, and shape. All radome designs must consider operating temperature, structural loads, vibration, wind, sand, hail, and rain.

For ground-based radomes, wind loading is the most important factor in the selection of a radome design. Other factors to be considered are humidity, blowing dust or sand, rain, ice, snow, and moisture buildup on the outer wall.⁷ The radome configuration and

mounting scheme will evolve from these environmental considerations and the electrical requirements. Typical surface-based structures include the metal space-frame,⁸ sandwich, monolithic-wall, and air-supported (inflatable) radomes.

Missile and aircraft radome designs are driven primarily by aerodynamic loads and the thermal environment. The selected configuration must be a trade between materials and shapes based on vehicle speed, trajectory, and electrical performance over the elevated temperatures produced by aerodynamic heating. Dielectric constant and loss tangent of radome materials can vary significantly with temperature. Non-uniform heating of the radome wall also produces changes in boresight-error performance.

Rain erosion is a severe problem in missile-radome operation and can be a limitation for aircraft radomes. The radome shape, material type, and velocity influence rain erosion. A sharp-nose radome is less susceptible to damage than a blunted-nose radome because less impact energy is transferred to the structure. Rain erosion manifests itself as pitting of the radome surface and, in severe cases, leads to catastrophic failure. Degraded RF performance can also result from the change in wall thickness. Special rain-erosion coatings and metal tips are used to provide resistance to rain erosion.

Water absorbed by a radome increases the dielectric constant and loss tangent of the wall. Preferred radome materials do not readily absorb water or must be treated with protective coatings. Water absorption also reduces the strength of most composite materials.

Static electricity caused by air friction can present a serious shock hazard. Thin antistatic coatings are used to neutralize static charge by providing a conducting path to attached structures. Lightning strikes to aircraft are common, so metallic lightning-diverter strips are used to minimize structural damage to the radome. Diverter cause some increase in sidelobe levels.

53.2 RADOME WALL DESIGN

This section discusses the configuration of radome walls and plane-wave analysis methods for predicting transmission and reflection properties. The importance of radome shape is also introduced.

Electrical Analysis

The most basic problem in radome electrical design is development of a wall configuration that supports the antenna system requirements and is also compatible with the overall mechanical and environmental requirements. This requires a model for the RF transmission and reflection properties of the wall. From the beginnings of radome use, it was recognized that tractable models can be based on some treatment of the incident field as a plane wave and the radome wall as an infinite planar slab. Solutions for this configuration are straightforward and have long been applied in radome design and analysis.¹⁻²

The conditions for a plane wave incident on a uniform, isotropic dielectric slab are shown in Figure 53-3. The *incidence angle* (θ_i) is the angle between the vector pointing to the source ($-\hat{k}$) and the surface normal (\hat{n} .) The plane containing \hat{k} and \hat{n} is the *plane of incidence*. Polarization is referenced to the plane of incidence with components in this plane, called *parallel polarized*, and perpendicular to the plane, called *perpendicular polarized*. These references are significant because, for isotropic materials, parallel polarized incidence results in completely parallel-polarized reflection and transmission, and likewise for perpendicular polarization. Therefore, these polarizations are *eigen-polarizations* of an isotropic slab.

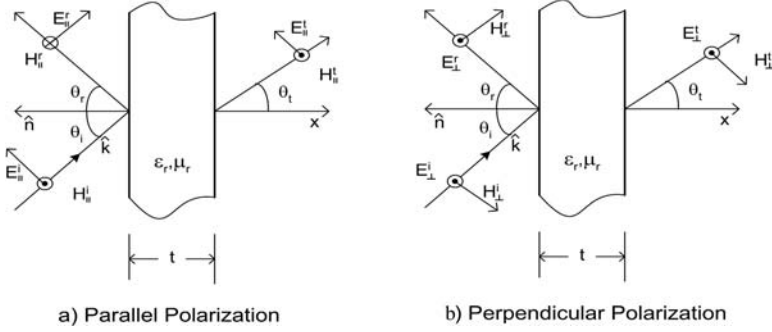


FIGURE 53-3 Conventions used for plane-wave propagation through a flat dielectric panel

The fields internal to the slab are a combination of forward- and backward-traveling waves produced by reflections between the interfaces. Boundary conditions require continuity of the tangential electric and magnetic at all material interfaces. Therefore, analysis can be based on the transverse field components and their relationship between adjacent boundaries. These fields and dielectric slabs are analogous to cascaded TEM transmission lines and the *ABCD transmission matrix* used for microwave circuits⁹ can also be used to describe the wall transmission and reflection properties. The transverse components of the total fields at the input side are made up of the incident and reflected fields, $E_i^i(1 + \Gamma)$ and $H_i^i(1 - \Gamma)$, and are related to the transverse components on the output side, E_t^i and H_t^i , by

$$\begin{bmatrix} E_i^i(1 + \Gamma) \\ H_i^i(1 - \Gamma) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} E_t^i \\ H_t^i \end{bmatrix} \quad (53-1)$$

Since wall interactions only depend on relative dielectric properties, all material parameters can be normalized to free space. For a slab with constitutive parameters, $\epsilon_r = \epsilon_r' - j\epsilon_r'' = \epsilon_r'(1 - j\tan\delta_e)$ and $\mu_r = \mu_r' - j\mu_r'' = \mu_r'(1 - j\tan\delta_m)$, the ABCD matrix is given by

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cosh(jk_x t) & Z_r \sinh(jk_x t) \\ Z_r^{-1} \sinh(jk_x t) & \cosh(jk_x t) \end{bmatrix} \quad (53-2)$$

where t is the slab thickness, λ_0 is the free space wavelength, and

$$Z_r = \mu_r / \sqrt{\mu_r \epsilon_r - \sin^2 \theta_i} \quad (\text{for perpendicular polarization})$$

$$Z_r = \sqrt{\mu_r \epsilon_r - \sin^2 \theta_i} / \epsilon_r \quad (\text{for parallel polarization})$$

$$k_x = \frac{2\pi}{\lambda_0} \sqrt{\mu_r \epsilon_r - \sin^2 \theta_i} \quad (\text{the propagation constant in the x-direction})$$

The slab transmission coefficient, $T = E^t / E^i$, and reflection coefficient, $\Gamma = E^r / E^i$, are given by

$$T = \frac{2}{A + B/Z_0 + CZ_0 + D} \quad (53-3a)$$

$$\Gamma = \frac{A + B/Z_0 - CZ_0 - D}{A + B/Z_0 + CZ_0 + D} \quad (53-3b)$$

where $Z_0 = 1/\cos\theta_i$ for perpendicular polarization and $Z_0 = \cos\theta_i$ for parallel polarization are the normalized transverse impedances for the incident and transmitted waves.

The phase of T is referenced to the output side of the slab. An important parameter for radome performance is the phase shift introduced by insertion of the slab into the propagation path. This is obtained by removing the free space phase shift from input to output to form the *insertion phase*, IP, given by

$$IP = \angle T + \frac{2\pi}{\lambda_0} t \cos\theta_i \quad (53-4)$$

For radome analysis, the phase of T is adjusted by the IP obtain the *insertion transmission coefficient*. Insertion phase can be a strong function of polarization and incidence angle, which for appreciably-curved radomes, leads to distortion of the transmitted wave-front and causes radome boresight errors and cross-polarization antenna responses.

The ABCD formulation is useful in analysis of multilayered structures because the output of one layer becomes the input to the adjacent layer. Therefore, the ABCD matrices of individual layers can be cascaded in a matrix product to obtain the overall transmission and reflection coefficients of all layers.

Wall Configurations

Radome wall structures are configured to address design requirements using available materials and construction techniques. This section describes the electrical characteristics of some basic configurations including homogeneous single-layer (monolithic), multilayer, space-frame, and grooved walls. Figure 53-4 shows the common wall types.

A *monolithic-wall* consists of a single constant-thickness slab of homogeneous dielectric material. The thickness of a monolithic wall radome may be selected to be electrically thin (less than approximately one-tenth wavelength) or an integer multiple of one-half wavelength in the dielectric. *Thin-wall* designs are applicable at lower frequencies where the permissible electrical thickness also provides adequate strength and rigidity, while multiple half-wave designs are used at higher frequencies where structural considerations prohibit the wall from being electrically thin and the operating bandwidth is sufficiently narrow (usually not more than 6–10%).

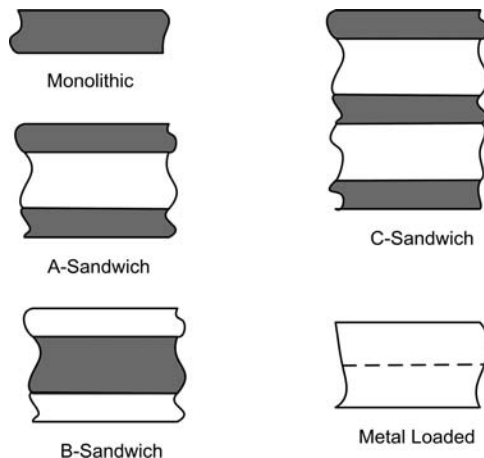


FIGURE 53-4 Radome wall-construction nomenclature

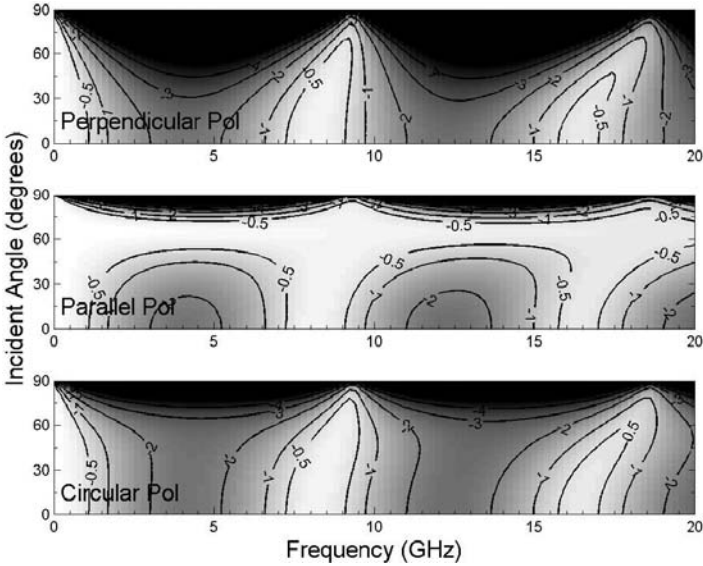


FIGURE 53-5 Transmission magnitude (dB) for a half-wave design at 9.0 GHz and $\theta_d = 60^\circ$ ($d = 8.6$ mm, $\epsilon_r = 4.5$, and $\tan \delta = 0.01$)

To illustrate characteristics of a monolithic wall, Figure 53-5 shows the transmission of a panel designed for 9 GHz and 60° incidence angle. Three polarization states are shown for performance over a 0–20 GHz frequency range and for 0– 90° incidence angles. High transmission is seen at very low frequencies, the design frequency, and twice the design frequency. Transmission for circular polarization is the complex average of parallel and perpendicular, $T_{CP} = (T_\perp + T_\parallel) / 2$, and is equal for right- and left-hand senses.

The low-frequency properties seen in Figure 53-5 apply to any electrically thin wall. These properties vary with wall thickness and dielectric, polarization, and incidence angle. Since there is low wall-absorption loss in low loss-tangent materials, the loss mechanism is mainly reflection. Thin-wall design curves for the maximum thickness required for a specified transmission level are presented by Kay.¹⁰

Transmission resonances occur in a monolithic wall at frequencies that produce integer multiples of a half wavelength thickness in the dielectric. The design thickness for these resonances, d , is given by

$$d = \frac{n\lambda_0}{2\sqrt{\epsilon_r - \sin^2 \theta_d}}, \quad n = 1, 2, 3, \dots \quad (53-5)$$

The integer n is the *order of the radome wall* with $n = 1$ being a *half-wave wall*. The incidence angle, θ_d , is the *design angle*. For lossless materials, reflection will be zero at this angle of incidence, transmission will be unity, and equal insertion phase will occur for perpendicular and parallel polarization. Thus, at the design angle and frequency, Figure 53-5 shows high transmission for all polarizations.

Transmission for parallel polarization is always greater or equal to that for perpendicular polarization. In addition, at the *Brewster angle*, θ_B , defined by

$$\theta_B = \tan^{-1} \sqrt{\epsilon_r} \quad (53-6)$$

lossless materials will have zero reflection and unity transmission for parallel polarization, independent of panel thickness and frequency. No similar condition exists for perpendicular polarization.

Monolithic radome walls are used in many ground-based and shipboard applications where weight is not a problem and simplicity of design and construction is desired. Half-wave and full-wave walls are also common in high-speed airborne radomes, however lighter sandwich structures are often used for lower-speed aircraft and missile applications. Complete transmission curves for monolithic wall designs can be found in the reference section.^{1,11}

Multilayer structures are used when a monolithic wall is unacceptable or does not provide adequate performance, especially in areas such as bandwidth, weight, or strength. Common multilayer configurations include the A sandwich, B sandwich, and C sandwich (Figure 53-4).

The *A sandwich* consists of three layers: two dense high-strength skins separated by a lower-density, lower-dielectric core material such as foam or honeycomb. This configuration can provide much higher strength for a given weight than a monolithic wall.

The design of an A sandwich usually begins by selection of the skin thicknesses and then the core thickness is chosen so that the reflected wave from the second skin cancels the reflected wave from the first skin at the desired frequency and incidence angle. Other than at normal incidence, the core thickness required for matching will be different for parallel and perpendicular polarization. Usually perpendicular matching is selected because good transmission for parallel is provided by the Brewster angle phenomena. Figure 53-6 illustrates transmission properties for an A sandwich designed for 9 GHz, 50° incidence, and perpendicular polarization.

Design formulas and graphical data for A-sandwich walls are presented in the reference section.^{1,11} For the symmetric A sandwich, the inner and outer skins are identical and are usually thin with respect to wavelength. For higher frequencies of operation where a half-wave wall would be structurally unacceptable, an A sandwich can be constructed

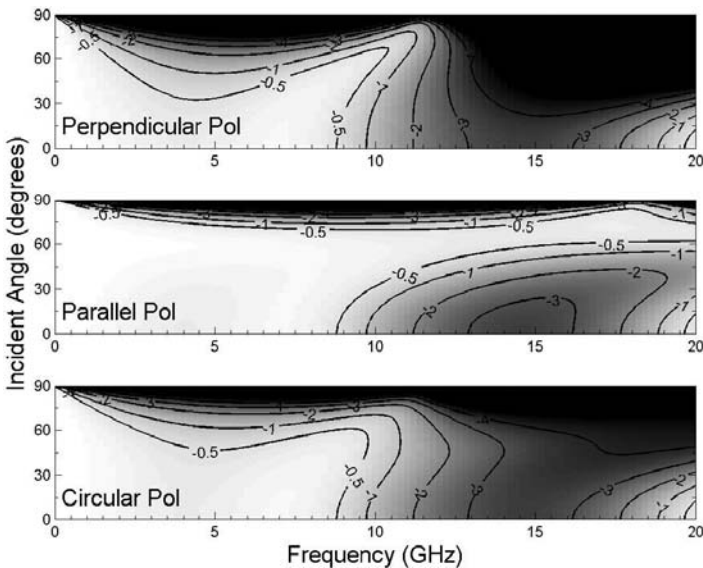


FIGURE 53-6 Transmission magnitude (dB) for a symmetric A sandwich designed for perpendicular polarization at an incidence angle of 50° and 9 GHz (skins: $t = 1.27$ mm, $\epsilon_r = 4.0$, $\tan \delta = 0.02$; core: $t = 7.75$ mm, $\epsilon_r = 1.2$, $\tan \delta = 0.005$)

using two half-wave tuned skins separated by an arbitrary-thickness of low-dielectric core, since matching will not require reflection from one skin to be cancelled by reflection from the other. A sandwich designs are used in many narrowband, blunt-nose radomes because the low incidence angles provide nearly uniform transmission properties. A-sandwich walls are also used in some streamlined missile radomes to obtain wideband performance.

The *B sandwich* has three layers, reversed to the A sandwich: a dense core material and two lower-density outer materials, which can serve as quarter-wave matching layers. A B sandwich can provide very wide bandwidth at high incident angles. But, because the low-density outer layers usually have poor structural or environmental properties, it is only used in applications where this is acceptable. When the outer layers are selected for quarter-wave matching, both the thickness and dielectric constant must be specified and these will be different for the parallel and perpendicular polarization.

The *C sandwich* consists of five layers, like two A sandwiches joined together. These additional layers may be used to provide greater strength and rigidity in the mechanical design and also give more degrees of freedom for the electrical design. For example, a good transmission match can be obtained independent of the insertion phase. This enables the design of a highly curved radome using varying layer thicknesses to obtain uniform amplitude and phase transmission characteristics and thereby create minimal impact on low sidelobe antenna performance.¹²

Multilayer structures may contain any number of dielectric layers needed to achieve desired properties such as high transmission over a broad frequency band or multiple bands, or to incorporate desired environmental and structural features. Initial design of multilayer radomes should always be based on consideration of the fundamental properties of the low-order structures. Optimization of the final design parameters can then be achieved using numerical techniques based on analysis of the cascaded wall layers.

Metal-loaded radomes consist of a combination of dielectric and conducting materials. The inclusion of conducting material in a radome wall provides an additional design component to achieve RF properties that cannot be obtained with all-dielectric structures. Conducting material may consist of thin metal films with a periodic pattern of features arranged in a lattice or an electrically-thick metal with apertures. A thin patterned-metal acts as a shunt admittance on the equivalent transmission-line model of the wall, whereas the thick wall with apertures acts as a waveguide transmission line and has waveguide properties such as low-frequency cutoff.¹³ An infinite planar array of such elements is commonly called a *frequency selective surface (FSS)* due to the frequency dependent properties not found in purely dielectric walls.

Metallic radomes offer the potential for higher operating temperatures, better rain-erosion resistance, high strength-to-weight ratio, decreased static-charge buildup, ease of handling, and reduced interference from out-of-band signals. Good transmission properties have been demonstrated on a metallic radome¹⁴ that uses slots in a thin metal outer skin.

Design and analysis of FSS structures can be based on an infinite planar periodic array (or grid) with plane-wave incidence, representation of the fields by a harmonic series (*Floquet modes*), and expansion of the induced currents or aperture fields using selected modal sets.¹⁵⁻¹⁸ Analysis can also be performed using the FDTD method.¹⁹ All FSSs exhibit resonant frequency behavior (band pass and band stop.) Operation of a radome with a FSS may be near a resonant frequency or well below resonance. Common elements used in the sub-resonance region include wire grids and capacitive platelets. Wire grids are useful because they act as shunt inductive circuit elements,¹³ which can be used to tune a thin-wall radome or the skins of an A sandwich.²⁰ Wire grids can also be used to increase the bandwidth of a half-wave design.¹¹ Arrays of electrically-small metal platelets act as shunt capacitive elements that can be embedded in a dielectric skin to increase its effective thickness.²⁰ Many types of metal-grid patterns have been described.^{11,15-16} FSS layers are also used in designs similar to radomes such as transmission and reflection polarizers, and dual-band subreflectors.¹⁵

A *space-frame radome* consists of a network of structural beams covered with a electrically thin dielectric material.^{8,11} It is often in the shape of a truncated sphere and is used in ground-based applications. The dielectric or metallic structural beams form triangular sections networked together to support the loads applied to the dielectric material. Although some antenna gain reduction is caused by the beams, their size and random orientation minimize undesired pattern distortion and boresight errors.

The *grooved radome* is a non-resonant structure consisting of a thick dielectric into which V-shaped grooves have been cut on both sides.²¹ The grooves on one side can be oriented 90° , with respect to those on the other side, to help cancel the anisotropic effects and equalize the transmission properties for orthogonal polarizations. Grooves provide a gradual transition from air to dielectric and back again, thus giving a continuous match to the incident wave analogous to pyramidal absorber materials. The groove depth determines the lowest frequency of operation, and the spacing the upper frequency limit.

Radome Shape

Radome shape and wall construction together comprise the radome configuration. Blunt radome shapes include the cylinder, the sphere, and combinations thereof. Since the incidence angles encountered in these shapes are generally less than 30° , such radomes have been called *normal-incidence radomes*² because they can be successfully designed by consideration of only the normal-incidence wall transmission properties, which are independent of polarization. Reflections can be a problem in blunt radomes since the reflected energy can be returned directly to the antenna and increase the VSWR in the feed system.²²

Streamlined shapes are used to reduce drag in nose radomes for airborne applications. Common streamlined shapes are the *ogive* (Figure 53-7), von Kármán, Sears-Haack, conical,

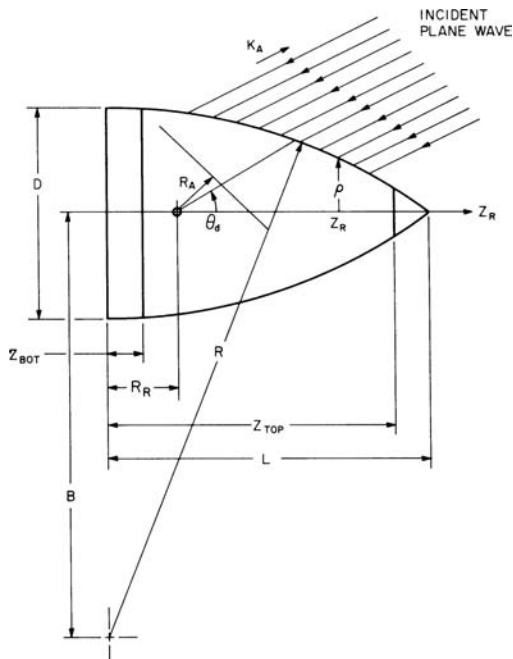


FIGURE 53-7 Antenna-radome geometry for a tangent ogive radome

ellipsoidal, power-series, and log-spiral.⁶ Incidence angles can range from 0 to 80° and higher, depending on the degree of streamlining which is often described by the radome length-to-diameter ratio called *fineness ratio*. Radome back reflection (VSWR) is not significant in streamlined radomes, but focusing effects, Lloyd's mirror effect, and trapped waves may have significant impact to electrical performance.¹

The superspheroid was developed as a parametric representation which can be used for a wide variety of radome shapes: both blunt and streamlined.²³

Radome shape affects both aerodynamic and electrical performance. In general, the more streamlined airborne shapes have less drag and are more resistance to rain damage; however, electrical performance is usually better for less streamlined shapes due to the lower incidence angles encountered. Sound design practice dictates that radome shape be optimized by considering both aerodynamic and electrical requirements.

53.3 ELECTRICAL ANALYSIS AND MODELING APPROACHES

This section describes methods that have been used to predict electrical performance of radomes and therefore can be used as design tools.

Antenna/Radome Interaction Phenomenology

The basic question in radome electrical design and analysis is how an antenna and radome will perform together as a system. This is a very difficult problem for most radomes because they are electrically large structures with complex geometry. Various quasi-optical techniques have historically been used for analysis. Recent advances in computational resources are now permitting more rigorous solutions as a boundary value problem. This section discusses the theoretical foundation for analysis and some techniques that have been applied.

Common to all radome analysis and modeling techniques are two primary issues:

1. How is RF transmission/reflection modeled at the radome wall?
2. How is near-field propagation modeled within the antenna/radome cavity?

One modeling technique is based on the equivalence principle.²⁴ In this approach, a transmitting antenna is considered and the fields transmitted to the radome interior are computed. Transmission through the wall is then calculated by a technique such as the plane wave/planar slab approximation to obtain the fields over the radome outer surface. Radiation to the far field is then determined from equivalent currents obtained from the tangential field components over the external surface. See Figure 53-8.

Another approach is based on the Lorentz reciprocity theorem.²⁴ In this method the received voltage is found by an integration that is performed over a closed surface, S_1 , surrounding the antenna. The integral requires the fields over S_1 that are produced by a source in the far field (with desired direction, frequency and polarization) and the fields produced by the antenna when transmitting. The voltage V_R received in each channel of a monopulse antenna is given by

$$V_R = c \oint\!\!\!\oint_{S_1} (\mathbf{E}_T \times \mathbf{H}_R - \mathbf{E}_R \times \mathbf{H}_T) \cdot \hat{\mathbf{n}} \, da \quad (53-7)$$

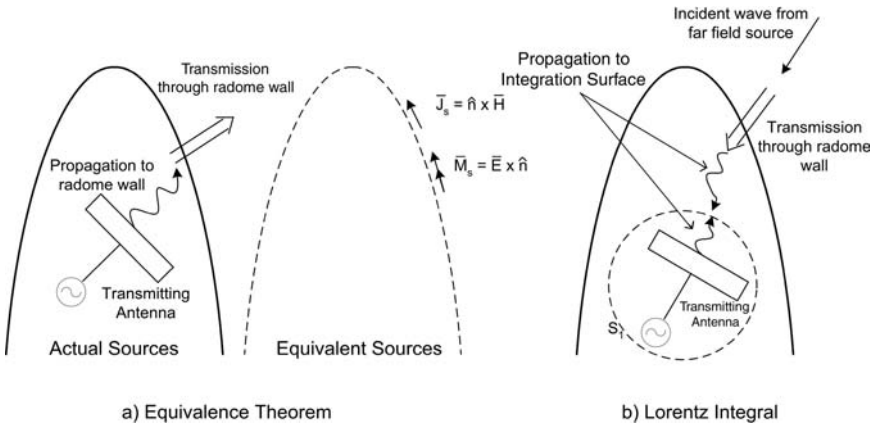


FIGURE 53-8 Use of equivalence and Lorentz reciprocity theorems to find antenna patterns in the presence of a radome

where

$\mathbf{E}_T, \mathbf{H}_T$ = antenna near fields when transmitting (sum or difference beams),
 $\mathbf{E}_R, \mathbf{H}_R$ = fields from the far-field source and incident on the closed surface S_1 ,
 $\hat{\mathbf{n}}$ = outward unit normal to surface S_1 , and
 C = a complex scale factor.

The integration surface may be chosen to coincide with the antenna aperture, the radome internal or external surface, or an arbitrary surface inside or outside the radome. It is the choice of the surface S_1 and the approximations used to find the fields over it which distinguishes the various methods of radome analysis.²⁵ Some techniques that have been used in modeling near-field propagation and wall transmission/reflection are summarized in Table 53-1 and are discussed next.

Radome analysis methods have paralleled the development of airborne radar systems. Silver² describes the geometrical-optics approaches taken up to 1949. By the mid 1960s radome design issues were well understood,¹ even though computational electromagnetics was still in its infancy. Principles identified by early researchers show great insight into antenna/radome interactions. Their work is well worth reading.

Perhaps the oldest analysis technique is referred to as ray tracing. It is usually described for a receiving antenna with the incident wave modeled by ray bundles. The radome wall is treated as being locally planar and modeled with the polarization-dependent, complex transmission coefficients for an infinite planar slab, lying in the tangent plane at the points

Table 53-1 Analysis Techniques for Modeling Transmission/Reflection at the Radome Wall and Propagation Within the Radome Cavity

Wall Transmission/Reflection	Propagation Within Radome Cavity
<ul style="list-style-type: none"> • Tangent Plane Approximation • Finite Element Method (FEM) • Finite-Difference Time-Domain (FDTD) • Integral Equation • Complex Ray 	<ul style="list-style-type: none"> • Ray-Tracing (diffraction ignored) • Near-Field Source Integration • Finite-Difference Time-Domain (FDTD) • Plane Wave Spectrum (PWS) • Finite Element Method (FEM)

of incidence of the impinging rays. The incidence angle and local polarization are determined by the surface normal and propagation vector. After applying transmission coefficients, the wave travels to the receiving aperture in the same direction as the incident field, thus ray tracing. Because diffraction effects due to changes in amplitude or phase are ignored, this is a geometric-optics approach. For electrically large systems, ray tracing is reasonably accurate in determining first order effects since only weak diffraction occurs between the radome and high-gain antenna. Ray tracing analysis is fast because extensive operations, such as matrix inversion, are not required.⁵

The limitations of ray tracing in modeling antenna near fields lead to techniques based on summation of distributed sources for predicting diffraction effects. Tricoles²⁶ formulated a three-dimensional method of radome analysis based on Shelkunoff's induction and equivalence theorems. Paris²⁷ applied these principles by integrating equivalent currents over the radome external surface in order to find the pattern of a horn antenna covered by a radome. Incident near-fields at the radome were found by a near-field summation and the tangent-plane approximation used for wall transmission. Tavis²⁸ describes a three-dimensional ray-tracing technique to find the fields on an equivalent aperture external to an axially symmetric radome.

Wu and Rudduck²⁹ developed a surface-integration method that uses an analytical plane-wave spectrum (PWS) representation to characterize the antenna near fields. This can be generalized for arbitrary apertures by use of the fast Fourier transform (FFT), which can also speed up the computer calculations.³⁰ Huddleston et al²⁵ have compared three-dimensional radome-analysis methods, which use a general formulation based on the Lorentz reciprocity theorem.

Siwiak et al³¹ applied the reaction theorem to the analysis of a tangent-ogive missile radome to determine boresight error. Their approach is based on surface integration of the antenna near fields and received fields carried out over the radome interior surface. The tangent plane approach was used to find the received fields at the radome interior and a model for one-bounce reflection from the radome far wall was based on ray tracing. A similar approach, but for the transmit case, has been developed for arbitrary shaped radomes.³²

Hayward et al³³ compared the accuracies of two methods of analysis in the cases of large and small radomes to show that ray tracing does not accurately predict wavefront distortion in small radomes. Burks et al³⁴ developed a ray-tracing analysis for large radomes which includes first-order reflections from the radome wall, and they demonstrated the correlation between wavefront distortion and boresight error including effects of polarization.

Gao and Felsen³⁵ presented a method of tracing complex rays through a tapered dielectric radome to find the near and far fields of the antenna. Orta et al³⁶ developed a Green's function technique to describe wall transmission characteristics independent of the excitation fields.

For most radomes, which are many wavelengths in size, integral-equation and finite-element methods (FEM) have been impractical due to the enormous computing resources required. However, the assumption of an axisymmetric shape greatly reduces the computational burden and forms the basis for a FEM method described by Gordon and Mittra³⁷ and the unimoment analysis of Kvam and Mei.³⁸ A volume integral approach for solution of two-dimensional radome problems has also been described by Sukharevsky et al.³⁹

A hybrid technique for electrically large axisymmetric radomes can be based on method-of-moments analysis of the wall in the nose region, tangent-plane approximations for the remaining wall, and analytical surface integration of equivalent currents to calculate antenna patterns.⁴⁰

A purely numerical approach for radome analysis can be implemented using the finite-difference time-domain (FDTD) technique, a method for solving the differential form of Maxwell's equations.⁴¹ FDTD uses the same model to account for both wall transmission/reflection and propagation. No added complexity is needed to account for tip and

bulkhead scattering. This technique is limited only by numerical accuracy and approximations needed to grid the radome and space, and the appropriate boundary conditions at the edges of the solution volume.

Effects on Antenna System Performance

Radomes will have some effect on every aspect of antenna performance and therefore affect all system parameters that rely on antenna characteristics. This includes changes in gain, sidelobe levels, beamwidth, null depth, and polarization. *Radome loss* is the difference in antenna gain with and without the radome present. Loss reduces system sensitivity, ERP, and link margin. Losses occur due to both impedance mismatch at the radome wall and dissipation within the wall. Loss can also be caused by increased antenna beamwidth due to amplitude and phase errors introduced by transmission through a curved wall and depolarization of the main beam.

Radome-induced changes in the amplitude and phase of antenna near fields usually increase far field sidelobe levels. Collimated reflection off the wall can also create a large *reflection sidelobe*, especially in highly streamlined radomes. Polarization of sidelobes will be changed by the radome and for some directions cross-polarized sidelobes will be higher than co-pol. Cross polarization is especially significant for the difference patterns used in monopulse tracking systems. Within the sum main-lobe, missile radomes can produce cross-polarized responses in the difference pattern that are as high as -20 dB with respect to sum. Furthermore, the slope of this cross-pol difference response is often opposite to that for co-pol, which will cause positive feedback in a tracking system and thus create a vulnerability to cross-pol jamming.

Boresight errors are caused by the asymmetric wavefront distortion of signals that pass through a radome and are received by angle-tracking antennas. This is an especially significant problem for homing missiles that use guidance based on estimates of target direction. Depending on the guidance system employed, boresight errors and boresight-error slopes affect missile performance in different ways. Consequently, radome specifications should be developed using consideration of overall tracking-system performance. Since it is impossible to obtain zero BSE at all scan angles (except for hemispherical radomes), angle-error compensation may be implemented in processing software based on measurements and lookup tables. If the table data is taken at ambient temperatures, there will be some uncertainty about its accuracy at elevated operating temperatures. Also, there may be some residual errors when receiving wave polarizations other than those used for making the compensation table.

Design Techniques

Radome design represents an enormous challenge due to the competing requirements from many disciplines. Material selection must be based on structural and thermal considerations. Shape is based on aerodynamic requirements and accommodation of antenna mounting and scanning. The wall configuration must support both electrical and structural goals. Radome design is always an iterative process based on understanding of the performance trades and component interactions.

The wall configuration is selected based on materials suitable for the thermal environment and a cross-section design that will support both the mechanical loads and the system center frequency and bandwidth. After selection of a baseline, one-dimensional thermal analysis can be performed to compute safety factors based on material strength and stress analysis of expected loads. Preliminary assessment of electrical performance can be based on flat panel analysis using incidence angles estimated by ray tracing.

Refinement of the design will occur as a result of these analyses. Optimization of the wall for minimum boresight error and gain loss over scan angle and frequency can be achieved by creating histograms of the incident angles and then tapering the wall thickness for best transmission and uniform insertion phase at the most likely incidence angles. Computer-aided analysis based on one of the techniques discussed previously will speed this process and provide performance predictions that can be used in system analysis. Two and three dimensional analysis of thermal and structural performance should be done in parallel with electrical optimization. Final design for a prototype will be based on engineering judgment with consideration of all analysis and verification then provided by testing.

53.4 RADOME MATERIALS

Electrical, structural, and thermal performance are primary factors in the selection of radome materials. Reliability, length of service, and cost can also be important considerations.

Ceramics

Ceramic materials that have been considered for radomes include aluminum oxide, beryllium oxide, boron nitride (pyrolytic and hot-pressed), magnesium aluminate (spinel), magnesium aluminum silicate (cordierite), magnesium oxide, silicon dioxide, sintered silica, and silicon nitride. The electrical properties of some ceramic radome materials are tabulated in Table 53-2. These electrical properties will vary depending on fabrication method, material density, and temperature.

Table 53-2 Ceramic Radome Materials (8.5×10^9 Hz, Ambient Temperature)

Material	Density, g/cm ³	Dielectric Constant	Loss Tangent
ALON, 9Al ₂ O ₃ -5AlN	3.69	9.28	0.0003
Aluminum oxide	3.32	7.85	0.0005
Alumina, hot-pressed	3.84	10.0	0.0005
Beryllium oxide	2.875	6.62	0.001
Boron nitride, hot-pressed	2.13	4.87	0.0005
Boron nitride, pyrolytic	2.14	5.12	0.0005
IRBAS	3.25	7.75	0.0012
Magnesium aluminate (spinel)	3.57	8.26	0.0005
Magnesium aluminum silicate (cordierite)	2.44	4.75	0.002
Magnesium oxide	3.30	9.72	0.0005
Pyroceram 9606	2.60	5.58	0.0008
Silicon dioxide	2.20	3.82	0.0005
Silica-fiber composite (AS-3DX)	1.63	2.90	0.004
Slip-cast fused silica	1.93	3.33	0.001
Silicon nitride	2.45	5.50	0.003

For high-speed homing missiles, currently used radome materials include Pyroceram 9606, slip-cast fused silica, and IRBAS (in-situ reinforced barium aluminosilicate). Silicon nitride also holds promise for radome applications due to its strength and hardness.

Radome materials for multimode missile seekers must have low transmission losses at some combination of microwave, millimeter-wave, optical, and infrared (IR) wavelengths. Materials that can be considered for this application include zinc sulfide, zinc selenide, magnesium fluoride, ALON, spinel, sapphire, yttria, diamond, and gallium arsenide.⁴²

Organic and Composites

Many thermoplastic and thermoset materials are suitable for radome use. Structural properties of resin systems can be enhanced by combination with a reinforcement material to create a composite. The electrical and structural properties of composite materials will depend on the matrix/reinforcement ratio, material purity, porosity, moisture content, temperature, fabrication process, and changes that occur during aging or environmental exposure. All these factors should be thoroughly investigated for any radome design. Some commonly used materials are listed in Table 53-3, and more comprehensive lists are found in the reference section.⁵⁻⁶

Thermoplastic materials are limited to applications involving modest temperatures, whereas laminates using polyimide or polybenzimidazole resins can withstand temperatures near 750°F. For higher temperatures, either ceramic materials or ablative coatings are usually required. Fiber-loaded Teflon (Duroid®) materials can be used at very high temperatures for a short period of time due to the temperature limiting properties of ablation. Duroid 5650 has been chosen as the best performer in a high-supersonic-rain environment.

Table 53-3 Organic Radome Materials (8.5×10^9 Hz, Ambient Temperature)

Material	Dielectric Constant	Loss Tangent
Thermoplastics		
Lexan	2.86	0.006
Teflon	2.10	0.0005
Noryl	2.58	0.005
Kydex	3.44	0.008
Composites		
Epoxy-E glass cloth	4.40	0.016
Polyester-E glass cloth	4.10	0.015
Polyester-quartz cloth	3.70	0.007
Polybutadiene	3.83	0.015
Fiberglass-polybenzimidazole	4.9	0.008
Quartz-reinforced bismaleimide	3.3	0.004
Quartz-reinforced cyanate-ester	3.3	0.005
Quartz-reinforced polyimide	3.2	0.008
Duroid 5650 (loaded Teflon)	2.65	0.003

Polyurethane and fluoroelastomer coatings are used for rain-erosion protection and, when slightly loaded with conductive fillers, also provide an antistatic coating. For adequate protection, the coating thickness is typically 0.01-0.02 inch (0.25-0.50 mm). For higher-speed missiles, rain-erosion coatings of polyurethane or neoprene are not used because of the higher temperatures involved. However ablative coatings such as Duroid and Avcoat offer a method to reduce rain erosion. Ablative coatings will change the wall thickness during flight and this must be considered in the radome electrical design. Protective coatings that cause beading of water have lower effects on electrical performance than those which produce continuous films.⁴³

53.5 FABRICATION AND TESTING

After design, a radome must be built and tested. This section summarizes common fabrication and testing techniques.

Fabrication Methods

Radome fabrication methods depend on the selected radome materials, and the final selection of materials is often based on the availability of a reliable and affordable manufacturing method. Construction methods for organic radomes include wet lay up, vacuum bag, autoclave, pressure bag, matched die, filament winding, resin transfer, foam in place, and injection molding.

Ceramic radomes have been constructed using slip casting, cold pressing, injection molding, hot pressing, glass forming, and flame spraying.⁶ Machining is often required for both organic and inorganic radomes to provide the necessary tolerances on wall thickness and interface dimensions.

Mechanical and electrical tests are performed throughout production. Before fabrication, tests are performed on specimens of the candidate radome structures to ensure that the required mechanical and electrical properties are reproducible within the required tolerances. During fabrication, quality-control tests are carried out to ensure correct shape, size, dielectric constant, loss tangent, and wall thickness. After fabrication, laboratory and operational tests are performed to validate radome performance in the field.

Electrical Test

Electrical testing of radomes includes measurements of complex permittivity of radome materials and electrical characteristics of the complete antenna-radome combination. Dielectric constant and loss tangent of material samples can be determined at microwave frequencies and high temperatures by using free-space transmission,⁴⁴ Fabry-Perot resonator,⁴⁵ and waveguide transmission or reflection techniques.⁴⁶ Accuracy is enhanced by use of modern network analyzers.⁴⁷

Radome electrical testing includes measurement of transmission efficiency, boresight error, boresight-error slope, beam deflection, radome reflection, antenna-pattern distortion, and depolarization as functions of antenna-radome orientation (scan) and frequency. Standard antenna range equipment, with the addition of a radome-test fixture,⁴⁸ can be used to measure pattern distortion, transmission efficiency, reflection (VSWR), and depolarization. Boresight-error measurements are more demanding because the indicated source direction must be accurately measured for varying antenna/radome geometry. Ways for

doing this include moving the source antenna on a xy-positioner in an automated closed-loop and using the indicated-angle error voltages in an open loop.⁴⁸

Mechanical and Aero/Thermal Test

Mechanical testing includes measurements of physical dimensions and testing of the structural, thermal, and rain-erosion properties of the complete radome. Ultrasonic and microwave-interferometer techniques can be used to measure wall thickness. Structural testing involves static loading of the radome, using mechanical, pressure, or vacuum methods in a manner simulating operational conditions. This must often be done at design temperatures. High-temperature and thermal-shock testing can be carried out by using oxyacetylene torches, quartz lamps, rocket motors, ramjet engines, solar furnaces,⁴⁹ and supersonic wind tunnels.⁵⁰⁻⁵¹ Rain-erosion testing has been carried out by using rocket sleds,⁵² rotating arms, shotgun lead pellets, and single and multiple impact waterjet apparatus.^{6,42} While highly desirable, it is extremely difficult to combine electrical and thermal tests, due to problems such as the reverberant RF environment in a wind tunnel, dynamics of a rocket sled test, and non-uniform heating of a solar furnace.

During radome development, computer-aided analyses should be used to predict the temperatures and mechanical stresses produced by aerodynamic loading and heating as functions of trajectory, altitude profile, velocity, and maneuvers of the vehicle. Parameters to be calculated include thermal-induced stresses, melting, dielectric changes, and maneuver loads versus trajectory time. Of particular interest is stress concentration in the tip region due to dissimilar materials and thermal expansion, and also bending stresses at the base of the radome due to aerodynamic pressure and maneuver. All radome test programs should be guided by the likely failure modes predicted by analysis.

REFERENCES

1. T. E. Tice (ed.), "Techniques for Airborne Radome Design," *AFAL-TR-66-391*, vol. 1 (December 1966): 6.
2. S. Silver (ed.), *Microwave Antenna Theory and Design* (New York: McGraw-Hill, 1949): 522-542.
3. J. D. Walton (ed.), "Techniques for Airborne Radome Design," *AFAL-TR-66-391*, vol. 2 (December 1966): 450.
4. K. D. Hill and J. D. Kelley, "Radome Design for High Power Microwave Systems," *Proc. 21st USAF Antenna R&D Symp.*, Air Force Avionics Laboratory and University of Illinois (October 1971): 1-49.
5. D. J. Kozakoff, *Analysis of Radome-Enclosed Antennas* (Boston: Artech House, 1997).
6. J. D. Walton, Jr. (ed.), *Radome Engineering Handbook: Design and Principles* (New York: Marcel Dekker, Inc., 1970).
7. I. Anderson, "Measurements of 20-GHz Transmission Through a Radome in Rain," *IEEE Trans. Antennas Propagat.*, vol. AP-23, no. 5 (September 1975): 619-622.
8. A. F. Kay, "Electrical Design of Metal Space Frame Radomes," *IEEE Trans. Antennas Propagat.*, vol. AP-13 (March 1965): 188-202.
9. D. M. Pozar, *Microwave Engineering* 3rd Ed. (Hoboken, NJ: John Wiley & Sons, 2005): 183-189.
10. A. F. Kay, "Radomes and Absorbers," *Antenna Engineering Handbook*, 1st Ed., Chap. 32, H. Jasik (ed.) (New York: McGraw-Hill, 1961).
11. V. J. Dicaudo, "Radomes," *Radar Handbook*, Chap. 14, M. I. Skolnik (ed.) (New York: McGraw-Hill, 1970).

12. B. Rulf, "Problems of Radome Design for Modern Airborne Radar," Part 1, *Microwave J.*, vol. 28, no. 1 (January 1985): 145–152; Part 2, vol. 28, no. 5 (May 1985): 265–271.
13. L. A. Robinson, "Electrical Properties of Metal-loaded Radomes," WADD Tech. Rept. 60-84, ASTIA-AD 249410 (February 1960).
14. E. L. Pelton and B. A. Munk, "A Streamlined Metallic Radome," *IEEE Trans. Antennas Propagat.*, vol. AP-22 (November 1974): 799–803.
15. T. K. Wu (ed.), *Frequency Selective Surface and Grid Array* (New York: John Wiley & Sons, Inc., 1995).
16. B. A. Munk, *Frequency Selective Surfaces: Theory and Design* (New York: John Wiley & Sons Inc., 2000).
17. T. Cwik, R. Mittra, K. C. Lang, and T. K. Wu, "Frequency Selective Screens," *IEEE Antennas Propagat. Soc. Newsletter* (April 1987): 5–10.
18. R. Mittra, C. H. Chan, and T. Cwik, "Techniques for Analyzing Frequency Selective Surfaces—A Review," *Proc. IEEE*, vol. 76, no. 12 (December 1988): 1593–1615.
19. J. G. Maloney and M. P. Kesler, "Analysis of Periodic Structures," *Advances in Computational Electrodynamics*, Chap. 6, A. Taflové (ed.) (Boston: Artech House, 1998).
20. R. H. J. Cary, "Radomes," *The Handbook of Antenna Design*, 2nd Ed., Chap. 14, A. W. Rudge, K. Milne, A. D. Oliver, and P. Knight, (eds.) (London: Peter Peregrinus, 1986).
21. D. G. Bodnar and H. L. Bassett, "Analysis of an Anisotropic Dielectric Radome," *IEEE Trans. Antennas Propagat.*, vol. AP-23 (November 1975): 841–846.
22. W. Cady, M. Karelitz, and L. Turner, *Radar Scanners and Radomes*, MIT Radiation Laboratory Series, vol. 26 (New York: McGraw-Hill, 1948).
23. P. L. Overfelt, "Superspheroids: A New Family of Radome Shapes," *IEEE Trans. Antennas Propagat.*, vol. AP-43 (February 1995): 215–220.
24. R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, Chap. 3 (New York: McGraw-Hill, 1961).
25. G. K. Huddleston, H. L. Bassett, and J. M. Newton, "Parametric Investigation of Radome Analysis Methods," *IEEE Antennas Propagat. Int. Symp. Dig.* (May 1978): 199–201.
26. G. P. Tricoles, "Radiation Patterns and Boresight Error of a Microwave Antenna Enclosed in an Axially Symmetric Dielectric Shell," *J. Opt. Soc. Am.*, vol. 54, no. 9 (September 1964): 1094–1101.
27. D. T. Paris, "Computer-Aided Radome Analysis," *IEEE Trans. Antennas Propagat.*, vol. AP-18, no. 1 (January 1970): 7–15.
28. M. Tavis, "A Three-Dimensional Ray Tracing Method for the Calculation of Radome Boresight Error and Antenna Pattern Distortion," Rep. TOR-0059 (56860)-2, Air Force Systems Command, May 1971.
29. D. C. F. Wu and R. C. Rudduck, "Plane Wave Spectrum-Surface Integration Technique for Radome Analysis," *IEEE Trans. Antennas Propagat.*, vol. AP-22, no. 3 (May 1974): 497–500.
30. E. B. Joy and G. K. Huddleston, "Radome Effects on Ground Mapping Radar," Cont. DAAH01-72-C-0598, AD-778 203/0, U.S. Army Missile Command, March 1973.
31. K. Siwiak, T. B. Dowling, and L. R. Lewis, "Boresight Error Induced by Missile Radomes," *IEEE Trans. Antennas Propagat.*, vol. AP-27, no. 6 (November 1979): 832–841.
32. J. A. Shifflett, "CADDRAD: A Physical Optics Radar/Radome Analysis Code for Arbitrary 3D Geometries," *IEEE Antennas and Propagation Magazine*, vol. 39, no. 6 (December 1997): 73–79.
33. R. A. Hayward, E. L. Rope, and G. P. Tricoles, "Accuracy of Two Methods for Numerical Analysis of Radome Electromagnetic Effects," *Proc. 14th EM Windows Symp.* (June 1978): 53–55.
34. D. G. Burks, E. R. Graf, and M. D. Fahey, "A High-Frequency Analysis of Radome-Induced Radar Pointing Error," *IEEE Trans. Antennas Propagat.*, vol. AP-30 (September 1982): 947–955.
35. X. J. Gao and L. P. Felsen, "Complex Ray Analysis of Beam Transmission Through Two-Dimensional Radomes," *IEEE Trans. Antennas Propagat.*, vol. AP-33, no. 9 (September 1985): 963–975.

36. R. Orta, R. Tascone, and R. Zich, "Performance Degradation of Dielectric Radome Covered Antennas," *IEEE Trans. Antennas Propagat.*, vol. 36, no. 12 (December 1988): 1707–1713.
37. R. K. Gordon and R. Mittra, "Finite Element Analysis of Axisymmetric Radomes," *IEEE Trans. Antennas Propagat.*, vol. 41, no. 7 (July 1993): 975–981.
38. T. M. Kvam and K. K. Mei, "The Internal Fields of a Layered Radome Excited by a Plane Wave," *IEEE Antennas Propagat. Int. Symp.* (June 1981): 608–611.
39. O. I. Sukharevsky, S. V. Kukobko, and A. Z. Sazonov, "Volume Integral Equation Analysis of a Two-Dimensional Radome With a Sharp Nose," *IEEE Trans. Antennas Propagat.*, vol. 53, no. 4 (April 2005): 1500–1506.
40. M. A. Abdel Moneum, Z. Shen, J. L. Volakis, and O. Graham, "Hybrid PO-MoM Analysis of Large Axi-Symmetric Radomes," *IEEE Trans. Antennas Propagat.*, vol. 49, no. 12 (December 2001): 1657–1666.
41. J. G. Maloney and G. S. Smith, "Modeling of Antennas," *Advances in Computational Electrodynamics*, A. Taflove (ed.) (Boston: Artech House, 1998): 453–456.
42. D. C. Harris, *Materials for Infrared Windows and Domes: Properties and Performance* (Bellingham, WA: SPIE Optical Engineering Press, 1999).
43. J. A. Effenberger, R. A. Strickland, and E. B. Joy, "The Effects of Rain on a Radome's Performance," *Microwave J.* (May 1986): 261–273; errata (August 1986): 145.
44. H. L. Bassett, "A Free-Space Focused Microwave System to Determine the Complex Permittivity of Materials to Temperatures Exceeding 2000°C," *Rev. Sci. Instrum.*, vol. 42, no. 2 (February 1971): 200–204.
45. K. H. Breeden and J. B. Langley, "Fabry-Perot Cavity for Dielectric Measurements," *Rev. Sci. Instrum.*, vol. 40, no. 9 (September 1969).
46. *Standard Test Methods for Complex Permittivity (Dielectric Constant) of Solid Electrical Insulating Materials at Microwave Frequencies and Temperatures to 1650°C*, ASTM Standard D2520, American Society for Testing and Materials.
47. *Basics of Measuring the Dielectric Properties of Materials*, Agilent Literature Number 5989-2589EN, Agilent Technologies Inc. (June 26, 2006).
48. T. J. Lyon, "Radome Measurements," *Microwave Antenna Measurements*, 3rd ed., Chap. 12, J. S. Hollis, T. J. Lyon, and L. Clayton, Jr. (eds.), (Atlanta: Scientific-Atlanta, Inc., 1985).
49. R. K. Frazer, "Duplication of Radome Aerodynamic Heating Using the Central Receiver Test Facility Solar Furnace," *Proc. 15th EM Windows Symp.*, (June 1980): 112–116.
50. D. J. Kozakoff, "Aerodynamically Heated Radome Measurement Techniques," Georgia Tech. Rep. A-2945, Cont. CFSI 4-81, F40600-81-0003, September 1981.
51. J. D. Walton, Jr., and H. L. Bassett, "A Program to Measure Radome/Antenna Patterns at High Temperature," Georgia Tech. Rep., Cont. DAAH01-73-C-0614, U.S. Army Missile Command, March 1974.
52. K. N. Letson et al., "Rain Erosion and Aerothermal Sled Test Results on Radome Materials," *Proc. 14th EM Windows Symp.* (June 1978): 109–116.

Chapter 54

Propagation

Joel T. Johnson

The Ohio State University

CONTENTS

54.1	INTRODUCTION AND OVERVIEW OF PROPAGATION EFFECTS	54-2
54.2	FREE SPACE PROPAGATION.....	54-3
54.3	ATMOSPHERIC REFRACTION.....	54-4
54.4	ATMOSPHERIC GAS AND RAIN ATTENUATION	54-6
54.5	EMPIRICAL PATH LOSS MODELS FOR LINE-OF-SIGHT LINKS.....	54-11
54.6	FADING AND MULTIPATH	54-14
54.7	IONOSPHERIC EFFECTS AT MICROWAVE FREQUENCIES	54-16
54.8	IONOSPHERIC AND GROUND WAVE EFFECTS AT FREQUENCIES LESS THAN 50 MHZ	54-17

54.1 INTRODUCTION AND OVERVIEW OF PROPAGATION EFFECTS

An antenna is a component in an electromagnetic system that provides an interface between the transmitter or receiver and the external environment. The effect of the external environment on electromagnetic propagation also is an important component of the electromagnetic system, and often plays a strong role in the design of the overall system as well as the antenna components. Because there are numerous environments in which antennas are employed (for example, underground, inside buildings or tunnels, inside the human body, and so on), a complete discussion of propagation effects would include a wide variety of such environments and the propagation behaviors unique to each. In addition, the physical means by which electromagnetic waves are transmitted from one location to another (called propagation “mechanisms”) are highly dependent on the electromagnetic wave frequency. Again a complete discussion would consider frequencies ranging from as low as the sub-kHz range up to optical frequencies or even beyond. These facts make clear that EM wave propagation in general is an immense subject, within which numerous topics and subdisciplines are contained. However, given the focus of the current volume on antenna design primarily at microwave frequencies (for example, around 300 MHz to 100 GHz), the discussion in this chapter is focused on microwave propagation effects in the Earth’s troposphere, as this is the environment in which many antenna applications occur. Some discussion of propagation effects at frequencies as low as 1 MHz in the Earth environment is also provided. While “reflections” by the Earth’s ionosphere are not a useful propagation mechanism at frequencies greater than around 50 MHz, ionospheric effects remain important in some applications up to frequencies as high as 6 GHz, as will be discussed in Section 54.7.

If sufficient antenna gain exists so that illumination of any obstacles (including the Earth’s surface) is avoided, the propagation environment becomes the Earth’s atmosphere and any weather-related features contained within. While this environment is reasonably approximated as free space in many situations, propagation effects including gaseous and rain attenuation, refraction, multipath fading, and atmospheric scintillations are present and can be determining factors in system design in some cases. If obstacles are illuminated, multipath fading becomes more prominent, and the diffraction mechanism becomes important if the line-of-sight path is obstructed. This chapter discusses propagation in free space, effects related to atmospheric refraction, gas and rain attenuation, empirical models for path loss on terrestrial links, basic multipath fading models, and ionospheric effects. References to more detailed information on these and other subjects are provided throughout.

It is important at this point to clarify the possible interaction of the propagation environment with an antenna’s properties. By coupling an electronic system to the outside world, the antenna becomes sensitive to its environment; therefore, properties such as its radiation pattern, impedance, and efficiency are strongly influenced by the location and surroundings of the antenna. It is important that the antenna designer consider these effects when developing a design in order to ensure that the desired antenna functionality is achieved when the antenna is placed in its operating location. For mobile systems, the antenna’s environment will change frequently, so that consideration of antenna functionality in a range of possible surroundings is desirable. While the distinction between propagation contributions and environmental effects on antenna properties in a given measurement can sometimes be difficult to make, most practicing engineers would describe the influence of structures (obstacles, water, and so on) located very close to the antenna as influencing the antenna rather than the propagation environment. This chapter excludes discussion of such environmental influences on antenna properties and focuses

instead on the typical meaning of propagation effects, which involve the contributions of the environment outside the very near region of the transmit or receive antennas. One related antenna parameter discussed elsewhere in this volume is the antenna noise temperature, which is also influenced by the external environment. See Chapter 50 (by Burnside et al) for further discussion.

While references for specific topics will be provided, a general resource for information on radio wave propagation is the Radio Communication Sector of the International Telecommunication Union (ITU-R).¹ The Recommendations series of the ITU-R describes methods for predicting electromagnetic wave propagation in a variety of situations; the methods presented have been agreed upon internationally following a detailed evaluation process, and continue to be revised as improved models are developed. The reader is recommended to examine the ITU-R web site to determine if further improvements to the Recommendations cited (current as of 2007) are available.

54.2 FREE SPACE PROPAGATION

If it is assumed that both transmit and receive antennas operate in a medium composed entirely of vacuum, the Friis formula (named for Harald Friis of Bell Laboratories who first proposed it in 1946) describes the power received (P_R , Watts) at a receiving antenna at a distance R in the far-field of a transmitting antenna to which P_T Watts of power are input:

$$P_R = P_T G_T(\theta_T, \phi_T) G_R(\theta_R, \phi_R) \left(\frac{\lambda}{4\pi R} \right)^2 |\hat{a}_E \cdot \hat{a}_R|^2 \quad (54-1)$$

where λ is the electromagnetic wavelength utilized and $G_T(\theta_T, \phi_T)$ and $G_R(\theta_R, \phi_R)$ represent the transmit and receive antenna gain patterns, respectively, evaluated at the angles at which the other antenna is viewed. The polarization factor $|\hat{a}_E \cdot \hat{a}_R|$ accounts for any mismatch between the polarization of the electric field impinging upon the antenna (\hat{a}_E) and the polarization properties of the receive antenna pattern (described by \hat{a}_R). It is important to recognize in non free-space environments that \hat{a}_E may not be precisely the polarization produced by the transmitting antenna, due to potential depolarization effects (caused for example by Faraday rotation in the ionosphere, by rain, by multipath effects, or by other mechanisms) in the propagation environment. Losses due to impedance mismatch effects, at either the transmit or receive sides, are neglected in Eq. 54-1.

It is often convenient to recast Eq. 54-1 in terms of decibel notation. If the powers are represented in terms of decibels above one Watt (dB,w), and if units of km and MHz are used for distance and frequency, respectively, Eq. 54-1 becomes

$$P_{R,\text{dBw}} = P_{T,\text{dBw}} + G_{T,\text{dB}}(\theta_T, \phi_T) + G_{R,\text{dB}}(\theta_R, \phi_R) - 20 \log_{10} R_{\text{km}} - 20 \log_{10} f_{\text{MHz}} \\ + 20 \log_{10} |\hat{a}_E \cdot \hat{a}_R| - 32.44 \quad (54-2)$$

This form of the Friis equation can be useful for simple system calculations. Note that the Friis equation states that received power is inversely proportional to the square of both the distance between the transmitter and receiver and the frequency utilized, although any frequency dependence of the antenna gains and the polarization mismatch must also be included.

54.3 ATMOSPHERIC REFRACTION

For systems operating in the Earth's environment, an apparently unimpeded path between transmitter and receiver does not consist of free space, but rather contains the Earth's atmosphere and any weather related features therein. This environment can have several influences on propagation, including refraction, attenuation, polarization modification, and so on. Here the refraction mechanism is discussed.

First, because the index of refraction of Earth's atmosphere n , in the absence of rain, snow, or other precipitation, is usually very close to unity, it is commonplace to discuss instead the "refractivity," N , defined through

$$n = 1 + N \times 10^{-6} \quad (54-3)$$

The term "N-units" is assigned to denote refractivity values. An expression for N valid over most frequencies of interest² is

$$N = 77.6 \frac{P}{T} + 3.732 \times 10^5 \frac{e}{T^2} \quad (54-4)$$

where P is the atmospheric pressure in mbar (or hPa), T is the temperature in Kelvin, and e is the water vapor pressure in mbar. The two terms of Eq. 54-4 are called the dry and wet refractivities, respectively. The water vapor pressure included in the wet term may be expressed in terms of the relative humidity H (specified in percent) as

$$e = \frac{6.1121H}{100} \exp \left[\frac{17.502(T - 273.15)}{T - 32.18} \right] \quad (54-5)$$

or in terms of the water vapor density ρ (in grams of water per cubic meter of atmosphere) through

$$e = \frac{\rho T}{216.7} \quad (54-6)$$

Because the total pressure, temperature, and water vapor pressure can all vary significantly depending on atmospheric conditions, the atmosphere's index of refraction is not spatially uniform, and refraction occurs. The dependence of the associated propagation effects on meteorology, which remains difficult to forecast precisely, motivates a statistical treatment of atmospheric refraction as well as many other propagation phenomena. It is commonplace to encounter probability distribution functions of many propagation parameters; these functions specify the percent of time that a particular parameter is likely to either exceed or fall below a given value, and therefore can be used in planning, for example, the number of hours per year that a given system problem is expected to be encountered. Most distribution functions are obtained empirically through averages over an entire year, although results are sometimes presented for the "worst month" as well in order to highlight behaviors over shorter time periods.³ Such information for the refractivity of the atmosphere can be found in the references;² further discussion of statistical effects will be provided in Section 54.6.

Despite the dependence on meteorology, it is often reasonable to assume that the atmosphere's properties are uniform horizontally over a propagation path (for example, distances up to tens of km), but vertical variations in atmospheric properties are pronounced

and cannot be neglected. A global mean description of the index of refraction as a function of height h is recommended as²

$$n(h) = 1 + (N_0 \times 10^{-6}) \exp \left[-\frac{h}{h_0} \right] \approx 1 + (N_0 \times 10^{-6}) - (N_0 \times 10^{-6}) \frac{h}{h_0} = n(0) + \left. \frac{dn}{dh} \right|_{h=0} h \quad (54-7)$$

where the final two forms apply for $h \ll h_0$, and with global mean values of 315 for N_0 at sea level and 7.35 km for h_0 . There is also additional information on expected properties of these parameters in specific global regions.² Note that weather phenomena can cause significant deviations from the typical values, and even the form of equation (54-7) can be invalid in some situations (for example, in ducting environments as will be discussed).

The assumption of horizontal homogeneity results in a locally spherically symmetric environment, and the associated form of Snell's law of refraction is

$$n(0) \cos[\psi(0)] = n(h) \left(1 + \frac{h}{a_E} \right) \cos[\psi(h)] \quad (54-8)$$

where a_E is the radius of the Earth (≈ 6370 km), and $\psi(h)$ is the angle at height h made by the ray path with respect to the local horizontal. If it is assumed that these angles are always small (as for terrestrial paths) and the linear approximation of $n(h)$ is used, Snell's law can be approximated as

$$\psi^2(h) = \psi^2(0) + \frac{2h}{a_E} \left[1 + \frac{a_E}{n(0)} \frac{dn}{dh} \right] \approx \psi^2(0) + 3.14 \times 10^{-4} h_{\text{km}} \left(1 + \frac{dN}{dh} \left[\frac{1}{157} \right] \right) \quad (54-9)$$

with the refractivity gradient $\frac{dN}{dh}$ in N-units/km and $\psi(h)$ in radians. This equation makes the two standard refraction phenomena apparent. First, if the refractivity gradient vanishes (i.e., $\frac{dN}{dh} = 0$), the ray angle nevertheless increases with altitude due to the changing horizontal direction associated with the curvature of the Earth.

The presence of a negative refractivity gradient (as is typical) decreases the rate at which the ray angle increases, with ray angles remaining constant or decreasing only for refractivity gradients more negative than 157 N-units/km. With such large negative refractivity gradients, ray paths are modified so that they return to the Earth's surface, from which they can be reflected, re-refracted back to Earth's surface, reflected again, and so on. This is called the ducting propagation mechanism, and can give rise to elevated received power levels due to its guided wave nature. The required rapid decreases in refractivity are not extremely unusual, especially over the ocean, but the linear approximation to the refractivity profile is usually not applicable for these cases, so that a ray-tracing computation using the non-approximated form of Snell's law and the true refractivity profile is required to estimate ray paths. Further discussion of the likelihood of ducting events is provided in Reference 2; note that ducting is generally not a reliable propagation mechanism due to the strong dependence on meteorology, but rather is usually a source of undesired interference.

In cases with moderate refractivity gradients where the linear approximation of the refractivity profile is reasonable over the ray path, Eq. 54-9 can be recast (again with $\psi(h)$ in radians) as

$$\psi^2(h) = \psi^2(0) + \frac{2h}{k_{\text{eff}} a_E} \quad (54-10)$$

where

$$k_{\text{eff}} = \left[1 + \frac{a_E}{n(0)} \frac{dn}{dh} \right]^{-1} = \left[1 + \frac{dN}{dh} \left[\frac{1}{157} \right] \right]^{-1} \quad (54-11)$$

is called the effective Earth radius multiplier, and where the refractivity gradient again is specified in N-units/km. Using this equation, propagation over a curved Earth including atmospheric refraction is modeled in terms of a propagation over an Earth with a modified radius of $k_{\text{eff}} a_E$ and with no atmospheric refraction (so that rays propagate in straight lines). Using the global mean values of N_0 and h_0 results in a k_{eff} value of 1.3755, more commonly approximated as 4/3. It is common practice in tropospheric radio wave propagation predictions to replace the real Earth surface with a sphere of radius $4/3 a_E$ and thereafter to ignore refraction effects. This approach is valid over a wide range of frequencies, but again, must be used with care as k_{eff} depends on meteorological conditions, with smaller values of k_{eff} (values as small as 2/3 are considered in many system planning stages) resulting in a higher degree of path blockage for line-of-sight systems. For ray paths propagating at larger angles or at higher altitudes, the spherical form of Snell's law and the true refractive index profile should be used to estimate ray-path properties.

In some situations, strong refractivity gradients in the atmosphere can produce scattering effects, leading to multipath interference effects; a brief discussion of the associated fading phenomena is provided in Section 54.6. Inhomogeneities in the atmosphere can also be used as scattering sources in a transhorizon propagation link; such tropospheric scatter communications systems are not discussed further here, but information can be found in the references.⁴ Variations in the atmospheric refractive index also cause scintillation phenomena, which are short-term variations in received signal power levels that can be interpreted as fading or in terms of variations in the angle of arrival along the transmission path. Scintillation effects become more important as the path length through the atmosphere increases (for example, at lower elevation angles on Earth-space paths), and at frequencies greater than 10 GHz. More information on the probability distribution function for scintillation effects is available in the references,⁵ along with further discussion of the expected temporal properties of scintillation-induced fading.⁶

54.4 ATMOSPHERIC GAS AND RAIN ATTENUATION

In addition to causing refraction, the atmosphere can also cause attenuation. At frequencies below 350 GHz this attenuation is due primarily to the atmospheric gases oxygen and water vapor and due to hydrometeors (liquid water or ice particles in the atmosphere including clouds, rain, snow, fog, hail, and so on). Of the latter, rain produces the largest propagation influence and therefore is the only hydrometeor contribution considered in detail here. Other hydrometeor effects may not be negligible, particularly at frequencies above 10 GHz, including those due to clouds or fog.⁷

Oxygen and water vapor both have molecular resonance frequencies below 350 GHz, and therefore can produce large attenuations for systems operating near these resonant frequencies. Both also contribute to a continuum absorption outside these resonant frequencies. The amount of attenuation encountered on a path is described through an integration of a local *specific attenuation* in dB/km over the path length. The specific attenuation encountered is described as a function of air temperature, pressure, and water vapor density, and therefore remains dependent on meteorology.⁸ Figure 54-1 is a plot of the specific attenuation contributions of oxygen and water vapor for a pressure of 1 atmosphere, temperature of 15 C,

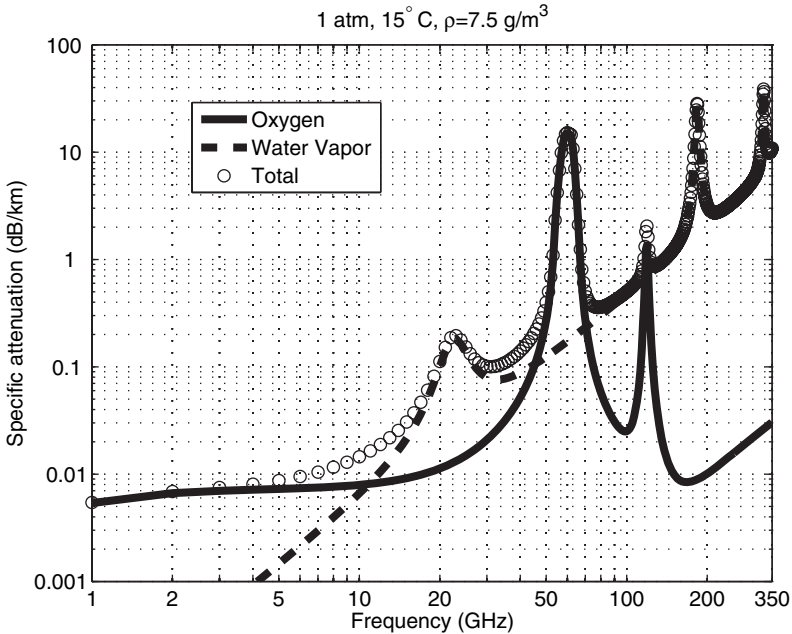


FIGURE 54-1 Specific attenuation due to atmospheric gases

and water vapor density of 7.5 grams per cubic meter (roughly 58% relative humidity). The curves show the water vapor resonant frequencies of 22.235, 183.31, and 325.15 GHz, as well as the broad oxygen resonance (composed of many individual resonances) near 60 GHz and the narrower oxygen resonance at 118.75 GHz. Note that near 60 GHz, atmospheric attenuation can be greater than 10 dB per km. Two methods for computing the specific attenuation are provided in the references: the first includes a complete summation over all known resonant lines, while the second describes a simplified curve fit to the results of the first method.⁸ The second approach was used in producing Figure 54-1; Figure 54-1 can be used to obtain a rough estimate of atmospheric specific attenuation for basic purposes; it is recommended to use the methods mentioned previously⁸ for more precise predictions.

Obtaining the total gaseous attenuation on a path requires an integration of the specific attenuation over the path length. For horizontal paths, it is usually reasonable to assume that atmospheric properties are uniform, so that the total attenuation is simply the specific attenuation (dB/km) times the length of the path (in km). However for Earth to space or other paths that travel through large heights above the surface, atmospheric properties and therefore the associated specific attenuation are not constant along the path, and an integration must be performed. Procedures are described for performing such an integration in the references, along with recommendations for sources of atmospheric meteorological information versus altitude, if such information is not available by local means.⁸

A simplified procedure for Earth-space paths is also available⁸ (for frequencies far from the resonant frequencies), in terms of an effective path length by which the specific attenuation can be multiplied to determine the attenuation for a path propagating through the entire atmosphere. Figure 54-2 illustrates results of this computation for zenith attenuation (for example, an Earth station communicating with a satellite directly overhead), again using (surface) parameter values of 1 atm pressure, 15 C, and 7.5 grams per cubic meter of

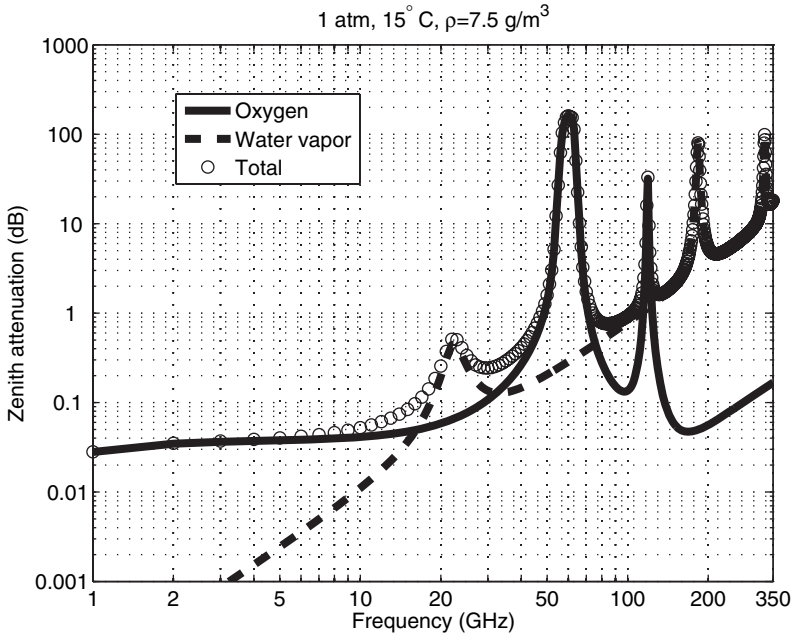


FIGURE 54-2 Zenith attenuation due to atmospheric gases

water vapor. The results continue to illustrate the strong attenuation at the resonant frequencies, and highlight the utility of the atmospheric “window” frequencies between resonances for Earth-space communications. The results of Figure 54-2 are highly dependent on the assumed altitude of the Earth station (assumed to be sea level in Figure 54-2); higher altitude stations would experience less attenuation due to the reduced atmosphere above them, as well as the likely reduced water vapor content at higher altitudes. Again Figure 54-2 should be used only for a rough estimation of total zenith attenuation, and not when more precise predictions are desired.⁸

For Earth-to-space paths where the satellite is viewed at elevation angles larger than five degrees, attenuation through the entire atmosphere can be estimated by dividing the zenith attenuation (in dB) by the sine of the satellite elevation angle. The larger values of attenuation obtained result due to the larger path lengths required to exit the atmosphere. This approach neglects the influence of refraction on the ray path: at smaller elevation angles, refraction becomes more important and a ray-tracing procedure must be used to ensure accurate calculations.

Numerous models, discussed extensively in the reference section,^{5,6,9,10} have been proposed for estimating rain attenuation. Rain attenuation is also described in terms of a specific attenuation that is integrated over the path length to obtain the total path attenuation due to rain. The specific attenuation due to rain, γ_R , has been found to be well approximated by the expression

$$\gamma_R = kR^\alpha \quad (54-12)$$

where R is the rain rate (in mm/hr) and k and α are functions of frequency and polarization; detailed expressions for k and α are provided in the references.¹¹ Figure 54-3 plots values

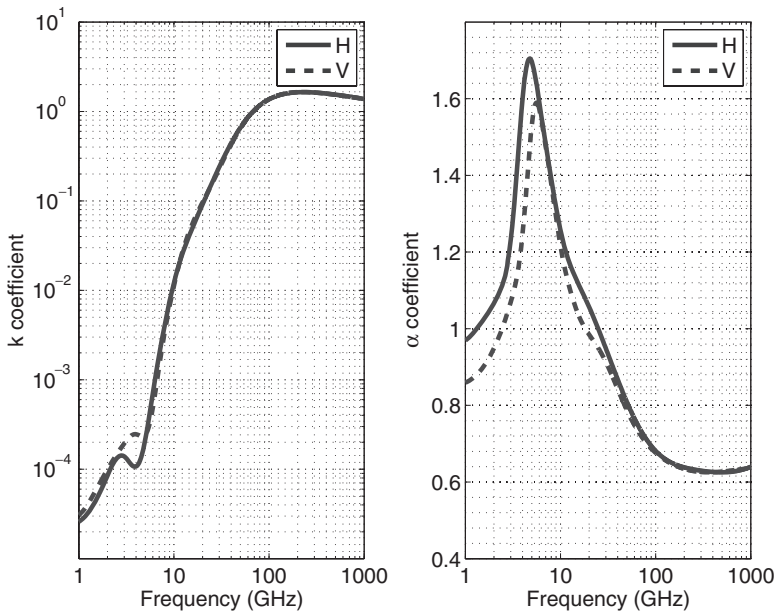


FIGURE 54-3 Coefficients k and α versus frequency for computation of rain specific attenuation

of k and α for horizontal (H) and vertical (V) polarizations: these plots can be used to obtain rough estimates for use in analysis of rain attenuation.

For polarizations other than horizontal or vertical, the values of k and α are given by

$$k = [k_H + k_V + (k_H - k_V) \cos^2 \theta \cos 2\tau] / 2 \quad (54-13)$$

$$\alpha = [k_H \alpha_H + k_V \alpha_V + (k_H \alpha_H - k_V \alpha_V) \cos^2 \theta \cos 2\tau] / (2k) \quad (54-14)$$

where θ is the path elevation angle and τ is the polarization tilt angle relative to horizontal polarization (45 degrees for circular polarization).

Because rain rates are not uniform horizontally or vertically, an integration of the varying specific attenuation over the path length is required to estimate total attenuation. The result of such an integration can be expressed as an effective path length by which the surface value of the specific attenuation should be multiplied to estimate total attenuation. The currently accepted ITU method for determining the effective path length on Earth-space paths⁵ is identical to that published originally¹² (although with typographical errors corrected¹³); for terrestrial links the current recommended procedure is also described in the reference section.¹⁰ The reader is referred to these references for the detailed equations.

These procedures allow attenuation to be computed given the surface value of the rain rate. Therefore a statistical model of the rain rate allows statistics of attenuation to be determined. While local measurements of rain statistics are the most desirable for use in such a procedure, there are also global methods for predicting average rain rates.^{6,9,10} The ITU models require knowledge only of the latitude and altitudes of the Earth stations and the corresponding rain rate exceeded 0.01% of the time. A probability scaling approach is

then used to map the attenuation exceeded 0.01% of the time to that exceeded at other time percentages. Note that different equations for this probability scaling are used in the case of terrestrial and Earth-space links. These values are useful in estimating the percent of time that rain attenuation may cause a system outage. In cases where some knowledge of rain attenuation is already available at a given location, equations are also provided for scaling such information to another frequency or polarization for which information is not available. Mitigation of rain attenuation through site diversity and the associated improvements in performance are described in the references.⁵

As an example of typical rain attenuation statistics, Figure 54-4 illustrates rain attenuation probability distribution functions for an Earth-space link at elevation angle 10 degrees, with the Earth station located at sea level, at latitude 40N, and with a rain rate of 49 mm/hr exceeded 0.01% of the time (these statistics roughly correspond to those for Columbus, OH, USA). The results illustrated are for circular polarization and for frequencies from 10 to 50 GHz. The results show that rain fading can cause extremely large attenuations, especially at higher frequencies. At lower frequencies, large fades are more rare (note that 0.01% of the time corresponds roughly to 1 hour per year, so the lower value on the horizontal axis corresponds roughly to six minutes per year). It is generally accepted that consideration of rain attenuation is generally required only for frequencies greater than 5 GHz. Due to the large attenuations that can be caused by rain attenuation, it is typically one of the dominant factors involved in planning system reliability, especially at frequencies greater than 10 GHz.

Both rain and ice particles can cause a change in the polarization state of a wave as it propagates, due to the non-spherical nature of these particles, which creates an effective anisotropic medium. Equations for predicting the cross-polarization discrimination (XPD)

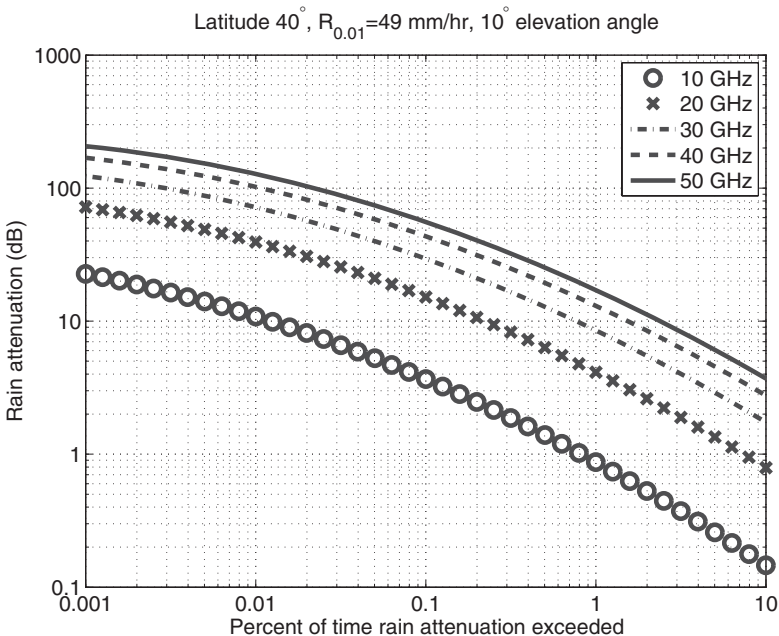


FIGURE 54-4 Rain attenuation example: attenuation probability distribution functions using the ITU-R model⁵

parameter (which describes the ratio of the received power in the transmitted polarization to that received in the orthogonal polarization) due to rain and ice particles are provided in the reference section.⁵ XPD values of 20 dB or less are possible on Earth-space links at higher frequencies; such values limit the extent to which frequency-reuse methods based on polarization diversity can be employed in a communications network.

54.5 EMPIRICAL PATH LOSS MODELS FOR LINE-OF-SIGHT LINKS

For propagation paths near the Earth's surface, the propagation environment is not simply free space, and the effects of the Earth's surface, terrain, and other obstacles must be considered. Two distinct approaches are available: in the first, it is assumed that knowledge of the local terrain is available, and efforts are made to develop a site-specific model based on known obstacles, buildings, and so on. In the second, it is assumed that no such information is available, and an empirical model for average propagation conditions is utilized. Numerous models have been developed in both of these cases, and recommendations continue to be refined as further information is obtained.

In the site-specific case, techniques ranging from standard rules-of-thumb combined with empirical information¹⁰ up to fully numerical electromagnetic simulations of the propagation environment have been reported. While the use of numerical techniques (either based on high frequency approximations and ray tracing or on direct solution of Maxwell's equations) is increasing, and commercial tools are becoming more widely available, this approach remains vastly more computationally expensive than the empirical approach, and therefore is not discussed further.

In the empirical site-specific approach, the local terrain (usually expressed in terms of height above sea level) is first considered and added to the sea-level profile of a modified curvature Earth (to account for atmospheric refraction as described in Section 54.3). Straight lines drawn between transmit and receive locations then provide information on the amount of path clearance above the local terrain or other obstacles. Should the path clearance be less than six-tenths of the first Fresnel zone distance at any point on the path, the path is considered obstructed and methods for predicting path loss in the obstructed case are used. Note that the first Fresnel zone distance in meters as a function of d_1 , the distance in km from the transmitter to a specific point on the propagation path, is defined as

$$F_1 = 17.3 \sqrt{\frac{d_1 d_2}{fd}} \quad (54-15)$$

where d_2 is the distance in km from the specific point to the receiver, d is the total path distance in km, and f is the frequency in GHz. Reference 10 provides an estimate of A_d , the average diffraction loss in dB for obstructed paths giving 15 dB or greater diffraction loss, as

$$A_d = -20 \frac{h}{F_1} + 10 \quad (54-16)$$

where h is the minimum path clearance above any significant obstacle, and F_1 is evaluated at the location where this minimum clearance is achieved. Further empirical information on expected statistics of diffraction fading is also available.¹⁰ Possible variations due to nonstandard atmospheric refractivity conditions in the effective Earth radius multiplier must also be considered in predicting expected minimum path clearances.

For non-obstructed paths, reflections from the Earth surface are a source of multipath interference. Reflections are typically obtained from relatively smooth and horizontal portions of the terrain, such as water surfaces or roads. Given a known reflection point, the time delay between the direct ray from transmitter to receiver and the ray following a path from the transmitter to the reflection point then to the receiver can be estimated. In this process, the reflection coefficient of the Earth's surface must be considered, but for long terrestrial paths it is typical to estimate this value conservatively as -1 due to the low grazing angles involved. Knowledge of the time (or phase) delay between direct and reflected rays then allows prediction of the antenna heights for which destructive interference is obtained. A step-by-step procedure for these analyses has been developed, and it recommends that slight modifications in antenna heights be used in order to reduce multipath when possible, while still maintaining acceptable path clearances.¹⁰ In cases where multipath cannot be avoided, multiple antennas at distinct heights can be used in a diversity approach.¹⁰

Figure 54-5 provides a simple illustration of a site-specific estimation of antenna heights for avoiding terrain obstructions. Propagation over a 36 km path is considered at a frequency of 600 MHz, with the transmitter at a height of 22 m above the local terrain on the left-most portion of the plot. Local terrain height values above sea level are illustrated in the plot, as well as these terrain values added to a $4/3$ radius Earth curvature plus six-tenths of the first Fresnel zone distance. The resulting straight ray drawn through the maximum obstacle for this path allows prediction of the receiver height (around 200 m above the local terrain in this case) in order to obtain non-obstructed propagation. The high antenna height required in this example would likely suggest a search for an alternate location due to the obstructed nature of this particular path.

An alternate approach to these site-specific methods relies entirely on empirical forecasts of propagation loss averaged over a wide range of conditions. This approach is more commonly applied for mobile systems, due to the continual changes in the path that occur. It is to be expected that these methods provide less accurate predictions for a particular

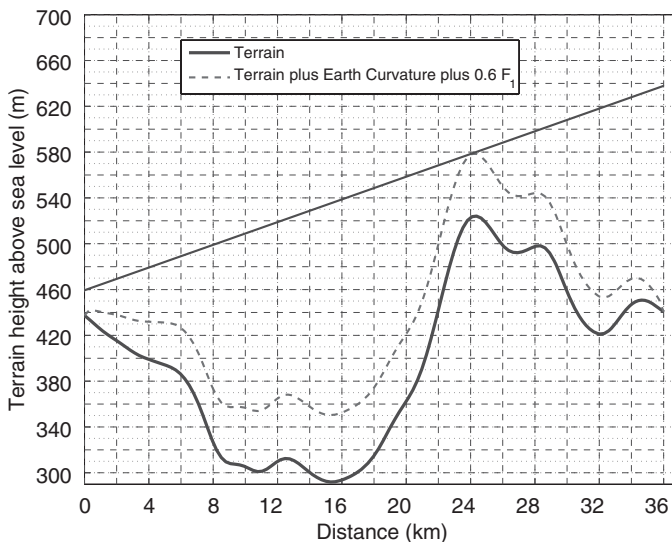


FIGURE 54-5 Site-specific example: estimation of antenna height to avoid obstruction due to terrain

path than those achievable with the site specific approach, but for systems that encounter a wide range of propagation paths, description of the average performance remains useful for predictions of expected path loss.

These models are typically expressed in terms of the path loss L_p in decibels, which in free space is identified as the terms $20 \log_{10} R_{\text{km}} + 20 \log_{10} f_{\text{MHz}} + 32.44$ from the Friis equation (Eq. 54-2). If propagation over a planar Earth is considered, both direct and reflected rays must be considered and added, including the phase delay between the two rays to obtain the total received field. If the factor $\frac{4\pi h_1 h_2}{\lambda d}$ is small (where h_1 and h_2 are the transmit and receive antenna heights above the ground, respectively, λ is the electromagnetic wavelength, and d is the path distance,) and the reflection coefficient of the planar Earth surface is taken as -1 , the path loss becomes

$$L_{p,\text{flat}} = 120 + 40 \log_{10} R_{\text{km}} - 20 \log_{10} h_1 - 20 \log_{10} h_2 \quad (54-17)$$

Unlike the free space case, this path loss increases as the fourth power of the range, and is independent of frequency.

Numerous empirical studies have shown that, while basic path loss behaviors for line-of-sight links are similar to those predicted by $L_{p,\text{flat}}$, improvements in this model are required for more accurate predictions. A plethora of revised path loss models have been proposed, typically using the form

$$L_p = L_0 + 10\gamma \log_{10} R_{\text{km}} + 10n \log_{10} f_{\text{MHz}} + L_1(h_1) + L_2(h_2) \quad (54-18)$$

where n and γ are empirically determined frequency and range scale factors, respectively.

One widely used model for forecasting path loss in land-mobile systems operating at frequencies from 150 to 1500 MHz is the Hata model.¹³ The Hata model is described as applicable for ranges from 1 to 20 km and assumes that antenna h_1 is a “base station” type antenna at a height ranging from 30 to 300 m, while h_2 is a mobile receiver antenna at a height ranging from 1 to 10 m. The specified path loss parameters are then

$$\begin{aligned} L_0 &= 69.55 && \text{Urban areas} \\ &= 64.15 - 2 \left[\log_{10} \left(\frac{f_{\text{MHz}}}{28} \right) \right]^2 && \text{Suburban areas} \\ &= 28.61 - 4.78 [\log_{10}(f_{\text{MHz}})]^2 + 18.33 \log_{10}(f_{\text{MHz}}) && \text{Open rural areas} \\ \gamma &= 4.49 \\ n &= 2.616 \\ L_1(h_1) &= \log_{10}(h_1)(-13.82 - 6.55 \log_{10} R_{\text{km}}) \end{aligned} \quad (54-19)$$

and

$$\begin{aligned} L_2(h_2) &= (1.56 \log_{10} f_{\text{MHz}} - 0.8) + h_2(0.7 - 1.1 \log_{10} f_{\text{MHz}}) && \text{Small to medium city} \\ L_2(h_2) &= 4.97 - 3.2 [\log_{10}(11.75h_2)]^2 && f_{\text{MHz}} > 300, \text{ large city} \\ L_2(h_2) &= 1.1 - 8.29 [\log_{10}(1.54h_2)]^2 && f_{\text{MHz}} < 300, \text{ large city} \end{aligned} \quad (54-20)$$

Note that the Hata model couples range and frequency dependencies into the antenna height and constant terms. Due to its basis on empirical data from Japan, the Hata model is often regarded as more applicable to very high density urban areas.

An extension of the Hata model to frequencies from 1.5 to 2 GHz is described in the reference section.¹⁴ The parameters of the modified model are identical to those of the original Hata model, with the exceptions

$$\begin{aligned} L_0 &= 46.3 && \text{Suburban/medium city} \\ L_0 &= 49.3 && \text{Large city} \\ n &= 3.39 \end{aligned} \quad (54-21)$$

Numerous other investigators have proposed a variety of empirical forms for average path loss,¹⁵ and in general it is found that differences among predictions of these models for supposedly similar conditions may be up to 30 dB or more! Therefore use of such empirical models should be restricted to conditions that are known to be similar to the datasets from which the models were derived.

Even given these limitations, the Hata model remains widely used for providing a first estimate of average path loss for land-mobile systems. Empirical models of the form of Eq. 54-18 have been widely proposed for a variety of propagation situations, including path losses inside buildings and rooms.

The references provide additional information on Earth-satellite mobile paths, including information on the contributions of attenuation in vegetation, diversity techniques, Earth-satellite propagation effects inside buildings, and fading properties and statistics.¹⁶

54.6 FADING AND MULTIPATH

The previous section primarily discussed methods for estimating the average path loss in either a fixed or mobile terrestrial link. However, even for a fixed link, variations in received power occur with time, often due to multipath contributions from obstacles in the vicinity of the propagation path. These variations are enhanced and more rapid for mobile systems. Because such variations are difficult to predict *a priori*, it is common to model multipath fading effects statistically, and to discuss probabilistic models for the received signals. Empirical studies have shown that the statistics of multipath fading can vary widely, depending on the propagation geometry and environment.^{5,6,10,16} Here the three most fundamental statistical models are reviewed in order to provide the reader sufficient background for further investigation.

Received power statistics for mobile systems are described here in terms of slow- and fast-fading components. The former results due to changes in the obstacles along a path as the mobile components move; this is a slow process because the dominant obstacles encountered vary slowly with time. Fast-fading effects are caused by multipath interference among a large number of received signal contributions; because these contributions are summed coherently, any small changes in the position of the receiver or the scatterers producing the multipath result in rapid changes in received power.

If the average path loss is known, it is typical to model slow fading losses in decibels L_S as a zero-mean Gaussian random variable with a specified standard deviation σ_{L_S} . This can be justified by writing the total slow fading loss as an integration of a slow fading specific attenuation over the path length; if it is assumed that the individual slow fading specific attenuations are random variables that are independent and identically distributed, the slow fading loss is then a Gaussian random variable by the central limit theorem. This analysis is reasonable only if there are many possible diffraction points that are encountered over time, as would be relevant for a mobile system. Including the effects of slow fading then simply involves modeling the received power (in dB,W) as a Gaussian random variable with a specified standard deviation. Link budgets can be assessed through use of the distribution

function of a Gaussian random variable. Values of σ_{L_S} ranging from 4 to 10 dB are typically used in propagation studies.

Fast-fading effects are more commonly treated by describing the received power in Watts (or the received voltage amplitude) as the random variable. If it is assumed that a dominant line-of-sight link exists (received power P_m watts neglecting multipath effects) to which multipath contributions from a large number of scatterers are added (received power P_f neglecting the line-of-sight term), the total received power random variable P follows Rician statistics, with a probability density function expressed in terms of an argument p of

$$f_p(p) = \frac{1}{P_f} e^{-\left(\frac{P_m+p}{P_f}\right)} I_0 \left(\frac{2\sqrt{P_m p}}{P_f} \right) \quad (54-22)$$

where I_0 is the modified Bessel function of the first kind and order zero. The associated distribution function is expressed in terms of the Marcum Q function and requires numerical evaluation. The mean value of the received power is given by $P_m + P_f$ and the standard deviation of received power by $\sqrt{P_f(P_f + 2P_m)}$. The ratio P_m/P_f is often called the Rician K-factor, and becomes larger as the line-of-sight component is more dominant.

Figure 54-6 plots Rician received power distribution functions for varying values of the K-factor (indicated in the legend), expressed in terms of the percent of time that the received power is less than X standard deviations below the mean received power. The sharp decreases for the curves at small values of K occur because the received power cannot become negative. Rician statistics are applicable when a primary path is received along with a large number of multipath contributions, and the approach is widely used in analyses of propagation channels.

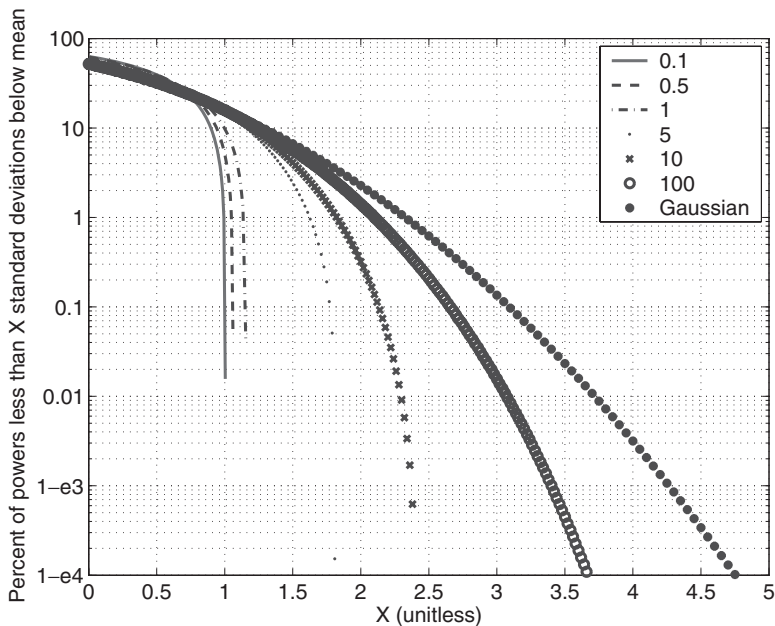


FIGURE 54-6 Rician distribution functions for specified K-factor

In the case that no line-of-sight link exists but a large number of multipath contributions remain, the received power follows Rayleigh statistics. The resulting probability density function is obtained from the Rician density by setting the line-of-sight term to zero; the mean and standard deviation of the received power are then both equal to P_f . The corresponding distribution function is simple, and states that the probability of the received power falling below p watts is $1 - e^{-p/P_f}$. For small values of p/P_f , this probability is simply p/P_f , implying that the power level exceeded q percent of the time in a Rayleigh channel is $(1 - q/100)P_f$.

Other statistical models of received powers have been proposed that attempt to simultaneously incorporate slow and fast fading effects, and numerous empirical distributions have been presented.^{10,15-17}

While the previous models provide information on the likelihood of a fade occurring at any time, no information on fade dynamics (such as fade duration) is contained. Empirical models for fade temporal properties have been described.^{10,16} Fading effects are of particular concern when they occur non-uniformly across the frequency bandwidth of the transmitted signal (known as “frequency selective” fading); this, of course, happens more frequently the wider the bandwidth of the transmission. Methods for mitigating multipath contributions through signal processing techniques are described in the communications systems literature.¹⁷

54.7 IONOSPHERIC EFFECTS AT MICROWAVE FREQUENCIES

The ionosphere is a region in the Earth’s atmosphere between altitudes of roughly 50 and 2000 km that contains a significant number of free electrons (due to ionization of atmospheric molecules primarily by solar radiation.) The presence of free electrons creates a plasma medium, and the resulting index of refraction neglecting the contributions of the Earth’s magnetic field is

$$n = \sqrt{1 - \left(\frac{f_{P,\text{MHz}}}{f_{\text{MHz}}} \right)^2} \quad (54-23)$$

where f_{MHz} is the frequency of operation in MHz and $f_{P,\text{MHz}} = 8.9788 \times 10^{-6} \sqrt{N}$ is the “plasma frequency” of the ionosphere, where N is the number of free electrons per cubic meter at a given location in the ionosphere (typical values range from roughly 10^9 to 10^{12} or higher). Eq. 54-23 shows that the index of refraction of a non-magnetized plasma takes on purely imaginary values for $f_{\text{MHz}} < f_{P,\text{MHz}}$ but approaches unity for frequencies $f_{\text{MHz}} \gg f_{P,\text{MHz}}$. Because ionospheric plasma frequencies are usually less than 10 MHz, it is clear that ionospheric effects are greatly reduced at frequencies greater than 0.5 GHz. However, for Earth-space paths that travel long distances through the ionosphere, even the slight deviations from unity that occur can cause important effects, although at frequencies greater than about 6 GHz nearly all ionospheric contributions become negligible. Because the ionosphere is created by solar radiation and also depends on the density of the atmosphere in order to supply molecules to be ionized, the number of electrons per cubic meter is a strong function of altitude, and also has major diurnal, seasonal, latitude, and solar-cycle (due to the approximate eleven year period of the sun’s sunspot number) variations. Spatial variations in the electron number density (especially with altitude) produce refraction and associated scintillation contributions.

The major ionospheric concerns for frequencies greater than 1 GHz caused by refraction effects are propagation delay, dispersion, and scintillation.¹⁸ All can be expressed in terms

of the total electron content (TEC) of the ionosphere on a given path, which is obtained by integrating N over the propagation path for a 1 square meter cross-sectional area (thereby obtaining the total number of electrons per square meter cross section). TEC values are often specified for paths integrated in the zenith direction, and values of 10^{16} and 10^{18} electrons per meter square are often used as representative of weak and strong ionospheric effects, respectively. A reference describes a procedure for predicting the TEC on a known propagation path.¹⁹ In terms of the TEC, an empirical expression for the average excess propagation delay (or group delay) in nsec relative to propagation in free space is

$$\Delta t, \text{nsec} = \frac{1.345}{f_{\text{GHz}}^2} \left(\frac{\text{TEC}}{10^{16}} \right) \quad (54-24)$$

Average time delays of 13.45 nsec are obtained, for example, at 1 GHz with a TEC of 10^{17} electrons per square meter. Because this delay is a known function of frequency, it is possible to estimate and remove its contributions by using measurements at two frequencies; this is the approach utilized in the Global Positioning System for correction of ionospheric time delays. Ionospheric dispersion and scintillation effects have also been discussed in detail.^{6,18-19}

When the effect of the Earth's magnetic field on the free electrons of the ionosphere is considered, it is found that the ionosphere is an anisotropic medium capable of producing changes in the polarization state of a wave as it propagates. As with any anisotropic medium, there are characteristic polarizations that remain unaltered by the medium; the characteristic polarizations of the ionosphere are a function of the angle between the direction of wave propagation and the Earth's magnetic field, but in many practical situations are approximately circular. Circular polarization is therefore preferred in most systems for space communications. For linear polarizations, the rotation of the linear polarization that results following propagation through the ionosphere is known as the Faraday rotation. An equation for the angle, Φ , in degrees through which a linear polarization is rotated following transit through the ionosphere is

$$\Phi = \frac{108}{f_{\text{GHz}}^2} \left(\frac{\text{TEC}}{10^{18}} \right) \quad (54-25)$$

Equation 54-25 assumes an average Earth magnetic field of 80 microteslas (a large value), but can be scaled linearly to other magnetic field values. As an example, a linear polarization is rotated 108 degrees at 1 GHz when the TEC is 10^{18} electrons per square meter. The reader is referred to the literature^{6,18-19} for a more detailed discussion of other ionospheric effects on Earth-space paths at microwave frequencies.

54.8 IONOSPHERE AND GROUND WAVE EFFECTS AT FREQUENCIES LESS THAN 50 MHZ

Though frequencies less than 50 MHz are not the focus of this chapter, a brief discussion of two major distinctions in propagation phenomena in this frequency range from those presented previously is provided. The first major change involves the ionosphere: at lower frequencies, significant changes in the ionospheric refractive index can be obtained, resulting in increased refraction of waves propagating in the ionosphere. This refraction mechanism can result in a ducting-like propagation mode in which rays broadcast upward from the ground surface can be refracted to turn back to the Earth's surface, often at large distances from the transmitter. This propagation mechanism is most effective in the HF frequency band (3-30 MHz) for long distance communications. At lower frequencies, the

ionosphere and Earth surface form the boundaries of a waveguide-like structure, so that signals can often be guided efficiently over long distances. Because ionospheric absorption tends to be reduced at higher frequencies, it is usually advantageous to operate at as high a frequency as is possible. However, given the reduced effect of the ionosphere at higher frequencies, eventually a frequency is reached at which there is insufficient refraction to return rays to the Earth's surface, and communication on a specific path is not possible. These properties result in a maximum usable frequency (MUF) for ionospheric reflections on a given path; the MUF depends on the geographical locations of the path endpoints as well as ionospheric conditions. Forecasts of MUF values and methods for their prediction are thus of importance in HF communications; numerous computer codes²⁰ have been developed for these purposes. Due to the dispersive nature of the ionospheric index of refraction (shown in Eq. 54-23), ionospheric communications over bandwidths larger than a few kHz are unusual, although frequency allocations of up to 2.7 MHz have been considered for advanced systems that can correct for propagation effects. Additional information on ionospheric propagation is available in the literature, and several recommendations in this area are available from the ITU-R.¹

The second mechanism important at frequencies less than 50 MHz is the groundwave, which is primarily important for vertically polarized transmissions with both transmitter and receiver at or near the ground. The groundwave is the dominant mechanism for local broadcasts of AM radio (around 1 MHz frequency) in the United States. Eq. 54-17, which describes propagation over a planar Earth, states that received powers vary as the fourth power of the range from the transmitter, and that the received power vanishes as both antennas approach the Earth's surface. The groundwave modifies these conclusions when it is present, and can produce a received power that varies as the reciprocal of the distance squared for a range of distances even with both antennas at or near the ground. Groundwaves are most significant when the imaginary part of the Earth's dielectric constant is high; this is typically the case at lower frequencies, and above approximately 50 MHz groundwave contributions are usually negligible at reasonable propagation distances. Ground wave field strengths can be reasonably predicted using analytical methods, based on a solution for dipole radiation in the presence of a finitely conducting Earth, combined with a theory of spherical Earth diffraction at larger distances. An overview of these theories and the associated prediction procedures and standardized plots of groundwave field strengths for simplifying system planning are available in the references.²¹⁻²² Note that at frequencies less than 50 MHz, typically both ionospheric and groundwave mechanisms can be important; in most cases, the groundwave dominates for local reception while the ionospheric term (or sky wave) dominates for long-distance reception. Variations in the ionosphere can also cause changes in these effects; in the MF band (300 kHz–3 MHz,) absorption in the lower ionosphere during the day reduces sky wave amplitudes. The return of the sky wave during night time hours typically forces a reduction in transmitted power for MF broadcast stations in order to avoid interference with other broadcasters.

REFERENCES

1. International Telecommunication Union, Radiocommunication Sector (ITU-R), <http://www.itu.int/ITU-R>.
2. ITU-R Recommendation P. 453-9, "The Radio Refractive Index: Its Formulation and Refractivity Data" (2003).
3. ITU-R Recommendation P. 581-2, "The Concept of Worst Month" (1990).
4. ITU-R Recommendation P. 617-1, "Propagation Prediction Techniques and Data Required for the Design of Trans-horizon Radio-relay Systems" (2002).

5. ITU-R Recommendation P. 618-8, "Propagation Data and Prediction Methods Required for the Design of Earth-space Telecommunication Systems" (2003).
6. L. J. Ippolito, *Propagation Effects Handbook for Satellite Systems Design*, 5th Ed. (Pasadena, CA: NASA Publication 1082, 1999).
7. ITU-R Recommendation P. 840-3, "Attenuation Due to Clouds and Fog" (1999).
8. ITU-R Recommendation P. 676-6, "Attenuation by Atmospheric Gases" (2005).
9. R.K. Crane, *Electromagnetic Wave Propagation Through Rain* (New York: Wiley-Interscience, 1996).
10. ITU-R Recommendation P. 530-11, "Propagation Data and Prediction Methods Required for the Design of Terrestrial Line-of-sight Systems" (2005).
11. ITU-R Recommendation P. 838-3, "Specific Attenuation Model for Rain for Use in Prediction Methods" (2005).
12. A. Dissanayake, J. Allnutt, and F. Haidara, "A Prediction Model That Combines Rain Attenuation and Other Propagation Impairments Along Earth-satellite Paths," *IEEE Trans. Ant. Prop.*, vol. 45 (1997): 1546–1558.
13. M. Hata, "Empirical Formula for Propagation Loss in Land-mobile Radio Services," *IEEE Trans. Veh. Tech.*, vol. VT-29 (1980): 317–325.
14. Kurner, T., "Propagation Models for Macrocells," Section 4.4 of COST Actions 231, *Digital Mobile Radio Toward Future Generation Systems: Final Report*, European Commission Technical Report EUR 18957 (1999): <http://www.lx.it.pt/cost231/>.
15. Lee, W. C. Y., *Mobile Communications Fundamentals*, 2nd Ed. (New York: Wiley, 1993).
16. J. Goldhirsh and W. Vogel, *Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems* (Austin, TX: NASA Publication 1274, 1998): <http://www.utexas.edu/research/mopro/>.
17. T. S. Rappaport, *Wireless Communications: Principles and Practice*, Chap. 7, 2nd Ed. (Upper Saddle River, NJ: Prentice Hall, 2001).
18. K. Davies and E. K. Smith, "Ionospheric Effects on Land Mobile Satellite Systems," (Pasadena, CA: NASA publication 1274, 2000): <http://www.utexas.edu/research/mopro/>.
19. ITU-R Recommendation P. 531-8, "Ionospheric Propagation Data and Prediction Methods Required for the Design of Satellite Services and Systems" (2005).
20. "High Frequency Propagation Models," Institute for Telecommunications Sciences web site (2006): <http://www.its.bldrdoc.gov/elbert/hf.html>.
21. J. Wait, "The Ancient and Modern History of Electromagnetic Groundwave Propagation," *IEEE Ant. Prop. Mag* (October 1998): 7-24
22. ITU-R Recommendation P. 368-8, "Groundwave Propagation Curves for Frequencies Between 10 kHz and 30 MHz," (2005).

Chapter 55

Materials and
Design Data

Donald G. Bodnar

MI Technologies

CONTENTS

55.1	PROPERTIES OF MATERIALS	55-2
55.2	ABSORBING MATERIALS	55-5
55.3	NONSOLID SURFACES	55-6
55.4	RADIO-FREQUENCY-BASED DESIGNATIONS.	55-13
55.5	ANTENNA-PATTERN CHARACTERISTICS	55-13
55.6	ELECTRICALLY SMALL HORNS	55-20

55.1 PROPERTIES OF MATERIALS

The design of many types of antennas, radomes, and impedance-matching devices requires the selection of appropriate dielectric materials. Strength, weight dielectric constant, loss tangent, and environmental resistance are the primary parameters of interest. Handbooks of mechanical properties are readily available¹⁻² and will not be reproduced here. Instead, emphasis will be on electric properties. The permittivity ϵ , permeability μ , and conductivity σ must be known at the operating frequency for a material to be completely specified electrically. For the dielectrics presented in this section, permeability equals that of free space ($4\pi \times 10^{-7}$ H/m), and permittivity is a complex number given by

$$\epsilon = \epsilon' - j\epsilon'' = \epsilon' (1 - j \tan \delta)$$

where $\epsilon'/\epsilon_0 = \epsilon_r$ and $\tan \delta = \sigma/(2\pi f\epsilon')$ are the dielectric constant and loss tangent, respectively, of the material, ϵ_0 is the permittivity of free space (8.854×10^{-12} F/m), and f is the operating frequency. For the metals discussed in this section, permeability equals that of free space, permittivity is unimportant, and skin depth equals $(\pi f \mu \sigma)^{-1/2}$.

Dielectrics

Foam dielectrics may be obtained in stock shapes or can be foamed in place (see Table 55-1). They are widely used as sealers for microwave components, spacers in twist reflectors and transreflectors, and low-cross-section target support. Specially loaded foams can be used to create a variable dielectric constant which is useful in a beam-collimation device such as a Luneburg lens. Foams and other low-density materials are typically used in low-power applications.

High-average-power applications (in the kilowatt range) rule out all known dielectrics except ceramics. In addition, high-speed missiles with their high surface and internal temperatures typically require ceramic radomes. Commonly used ceramics are listed in Table 55-1.

Plastics (Table 55-1) are widely used as structural members for antennas, feeds, and mounting structures. Conducting surfaces can be added by flame spraying, painting, and so on, to form reflectors. Sandwich construction techniques are used in radomes and polarizers. Here, two or more thin skins are stabilized against buckling with a lightweight core. The core material can be rigid foam (Styrofoam), foam-in-place foam (Eccofoam S), or a honeycomb material (Hexcell HRP). The skins can be epoxy fiberglass, Teflon, or Mylar. The thicknesses of the skins and core are chosen to make the sandwich transparent electrically. Careful control of tolerances, glue-line thickness, and materials can produce repeatably accurate devices. Photoetched printed-circuit boards are used in polarizers, spatial filters, and frequency-selective reflecting surfaces. Additional information on dielectric properties is given in Von Hippel.³

Metals

Copper, brass, and aluminum are the most important metals in antenna construction today. If weight is not of primary importance, brass and copper are used extensively. These two metals machine, solder, and plate so easily, with little or no special equipment, that they are favorites of the model shop and the production shop alike. Complicated assemblies can

TABLE 55-1 Dielectric Data

		Frequency f , Hz; Values of $\tan \delta$ Multiplied by 10^4				
		$f = 10^2$	$f = 10^4$	$f = 10^6$	$f = 10^8$	$f = 10^{10}$
Ceramics						
Alsimag 243	ϵ_r	6.30	6.28	6.22	6.10	5.76
	$\tan \delta$	12.5	4.0	3.7	3	8.5
Alsimag 393	ϵ_r	4.95	4.95	4.95	4.95	4.95
	$\tan \delta$	38	16	10	10	9.7
Aluminum oxide; Coors AI-200	ϵ_r	8.83	8.82	8.80	8.80	8.79
	$\tan \delta$	14	4.8	3.3	3.0	18
Beryllium oxide	ϵ_r	4.61	4.41	4.28	4.20
	$\tan \delta$	17	74	38	5
Magnesium oxide	ϵ_r	9.65	9.65	9.65	9.65
	$\tan \delta$	3	3	3	3
Porcelain, dry-process	ϵ_r	5.50	5.23	5.08	5.02	4.74
	$\tan \delta$	220	105	75	98	156
Silicon nitride	ϵ_r	5.50
	$\tan \delta$	30
Steatite 410	ϵ_r	5.77	5.77	5.77	5.77	5.7
	$\tan \delta$	55	16	7	6	22
Titania	ϵ_r	100	100	100	100	90
	$\tan \delta$	23	6.2	3	2.5	20
Glasses						
Fused quartz	ϵ_r	3.78	3.78	3.78	3.78	3.78
	$\tan \delta$	8.5	6	2	1	1
Fused silica	ϵ_r	3.78	3.78	3.78	3.78	3.78
	$\tan \delta$	6.6	1.1	0.4	0.3	1.7
E glass	ϵ_r	6.43	6.39	6.32	6.22	6.11
	$\tan \delta$	42	27	15	23	60
Low-density materials						
Eccofoam S	ϵ_r	1.05	1.18	1.47
	$\tan \delta$	20	40	70
Hexcel HRP, 1/4-4.5; honeycomb	ϵ_r	1.09
	$\tan \delta$	30

(Continued)

TABLE 55-1 Dielectric Data (Continued)

		Frequency f , Hz; Values of $\tan \delta$ Multiplied by 10^4				
		$f = 10^2$	$f = 10^4$	$f = 10^6$	$f = 10^8$	$f = 10^{10}$
Hexcel HRP, 1/4–4.5; honeycomb Styrofoam 103.7	ϵ_r	1.12
	$\tan \delta$	30
	ϵ_r	1.03	1.03	1.03	1.03
	$\tan \delta$	2	1	2	1.5
Plastics						
Bakelite	ϵ_r	8.2	6.5	5.4	4.4	3.52
	$\tan \delta$	1350	630	600	770	366
Duroid 5650	ϵ_r	2.65
	$\tan \delta$	30
Epoxy resin RN-48	ϵ_r	3.64	3.61	3.52	3.32	2.91
	$\tan \delta$	31	68	142	264	184
Fiberglass; laminated BK-174	ϵ_r	14.2	7.2	5.3	4.8	4.37
	$\tan \delta$	2500	1600	460	260	360
Lexan	ϵ_r	2.86
	$\tan \delta$	60
Plexiglas	ϵ_r	3.40	2.95	2.76	2.59
	$\tan \delta$	605	300	140	67
Polystyrene	ϵ_r	2.56	2.56	2.56	2.55	2.54
	$\tan \delta$	0.5	0.5	0.7	1	4.3
Rexolite 1422	ϵ_r	2.55	2.55	2.55	2.55	2.54
	$\tan \delta$	2.1	1	1.3	3.8	4.7
Polyethylene	ϵ_r	2.25	2.25	2.25	2.25	2.24
	$\tan \delta$	2	2	2	2	6.6
Polyvinyl chloride, W-176	ϵ_r	6.21	4.70	3.53	3.00
	$\tan \delta$	730	1070	720	500
Teflon	ϵ_r	2.1	2.1	2.1	2.1	2.08
	$\tan \delta$	5	3	2	2	3.7

Note: Properties are typical for a class of materials. Consult manufacturers for exact properties of their material.³

be built with practically no radio-frequency (RF) discontinuities due to contact between parts. Aluminum equals or surpasses both copper and brass on all counts except plating. Although special equipment is required to weld aluminum, the resulting structure is lighter in weight than brass or copper. Very complicated shapes such as flat-plate arrays can be fabricated by dip-brazing aluminum. Table 55-2 lists the conductivity of commonly used metals.

TABLE 55-2 Conductivity of Various Metals

Material	Conductivity, s/m at 20°C
Aluminum, commercial hard-drawn	3.54×10^7
Brass, yellow	1.56×10^7
Copper, annealed	5.80×10^7
Copper, beryllium	1.72×10^7
Gold, pure drawn	4.10×10^7
Iron, 99.98 percent pure	1.0×10^7
Iron, gray cast	$0.05 - 0.20 \times 10^7$
Steel	$0.5 - 1.0 \times 10^7$
Lead	0.48×10^7
Nickel	1.28×10^7
Silver, 99.8 percent pure	6.14×10^7
Tin	0.869×10^7
Tungsten, cold-worked	1.81×10^7
Zinc	1.174×10^7
Titanium	0.182×10^7

55.2 ABSORBING MATERIALS

Microwave absorbing material is commercially available in sheet, pyramidal, block, and rod form. Pyramidal absorber is commonly used to create an RF anechoic chamber for indoor-antenna range measurements since it offers the best performance (reflected energy 50 dB down at normal incidence). Pyramidal absorber is usually constructed by loading a spongelike rubber with graphite, and it is available in 2- by 2-ft (0.6- by 0.6-m) pieces ranging in thickness from 3 in (76 mm) to 12 ft (3.7 m). It is rated in decibels of incident to reflected energy at stated frequency, polarization, and incidence angle (see Table 55-3). Performance degrades if the pyramidal tips are broken or if the absorber is exposed to weather for extended periods.

Thicker absorber is typically used only in the specular regions of a chamber, in order to reduce cost. Staggering the thickness of the absorber according to a Chebyshev distribution⁴ can be used to improve the low frequency reflectivity.

Flat sheets of absorber are available when space is limited or when the absorber will be exposed to weather, wear, or high power. The reflected energy from sheet absorber is 15 to 25 dB greater than that of pyramidal absorber. Blocks and special shapes of absorber are available for waveguide-load applications, or they can be machined to the desired shape.

The previous types of absorbers are currently available commercially from Cuming Microwave Corporation, Avon, Massachusetts and TDK RF Solutions, Cedar Park, Texas.

TABLE 55-3 Reflection Coefficient of Typical Commercial Pyramidal Absorbing Material

Nominal Thickness, in	Reflection Coefficient, dB									
	120 MHz	200 MHz	300 MHz	500 MHz	1 GHz	S band	C band	X band	Ku band	K band
3	30	40	45	50
5	30	40	45	50	50
8	30	40	45	50	50	50
12	35	40	45	50	50	50
18	30	40	45	50	50	50	50
24	30	35	40	50	50	50	50	50
45	..	30	35	40	45	50	50	50	50	50
70	30	35	40	45	50	50	50	50	50	50
106	35	40	45	50	50	50	50	50	50	50
144	40	45	50	50	50	50	50	50	50	50

55.3 NONSOLID SURFACES

Antenna structures are sometimes made from nonsolid metal in order to make the structure lighter, to reduce wind loading, or for special RF purposes (for example, a frequency-selective reflector). Typical construction techniques employ metal wires, rods, strips, patches, perforated metals, and expanded mesh. How large the holes in the structure can be depends on how much transmitted or reflected loss can be tolerated.

Gratings

An array of identical parallel metal wires or rods can be used to reflect energy polarized parallel to the wires and, consequently, to shield the region on the other side of the wires for parallel polarization. The perpendicularly polarized energy is negligibly affected by the wires for most geometries. The parallel and perpendicularly polarized power-transmission coefficients are given, respectively, by⁵

$$T_{||} = 1 - \frac{1}{1 + \left(\frac{2S}{\lambda} \ln \frac{S}{\pi D} \right)^2}$$
$$T_{\perp} = 1 - \frac{\left(\frac{\pi^2 D^2}{2\lambda S} \right)^2}{1 + \left(\frac{\pi^2 D^2}{2\lambda S} \right)^2}$$

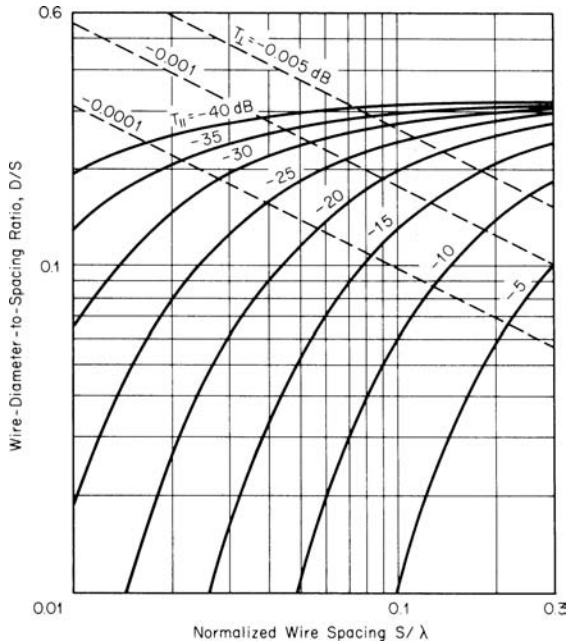


FIGURE 55-1 Transmission coefficient for parallel polarization and for perpendicular polarization for a grid of round wires of diameter D and center-to-center spacing S

Figure 55-1 is a plot of these equations, where the grating has wires of diameter D , center-to-center spacing of S , and wavelength λ . These curves are derived from Lamb's work⁶ and agree well with Mumford's nomograph.⁷ The grating can be envisioned as waveguides below cutoff. Thus to be effective, (1) the electric field of the incident wave must lie in the plane determined by the axis of the grating element and the direction of propagation, and (2) S must be less than $\lambda/(1 - \sin \theta)$, where θ is the angle between the direction of propagation and the normal to the axis of the grating element. Larger spacings result in undesirable secondary reflection lobes.⁶

The round wires may be replaced by thin metal strips of width W perpendicular to the direction of propagation by using metal bars or printed-circuit-board construction techniques. The transmission coefficients for the thin broadside strips are given in Figure 55-2, which are plots of the equations⁵

$$T_{||} = 1 - \frac{1}{1 + \left(\frac{2S}{\lambda} \ln \sin \frac{\pi W}{2S} \right)^2}$$

$$T_{\perp} = 1 - \frac{\left(\frac{2S}{\lambda} \ln \cos \frac{\pi W}{2S} \right)^2}{1 + \left(\frac{2S}{\lambda} \ln \cos \frac{\pi W}{2S} \right)^2}$$

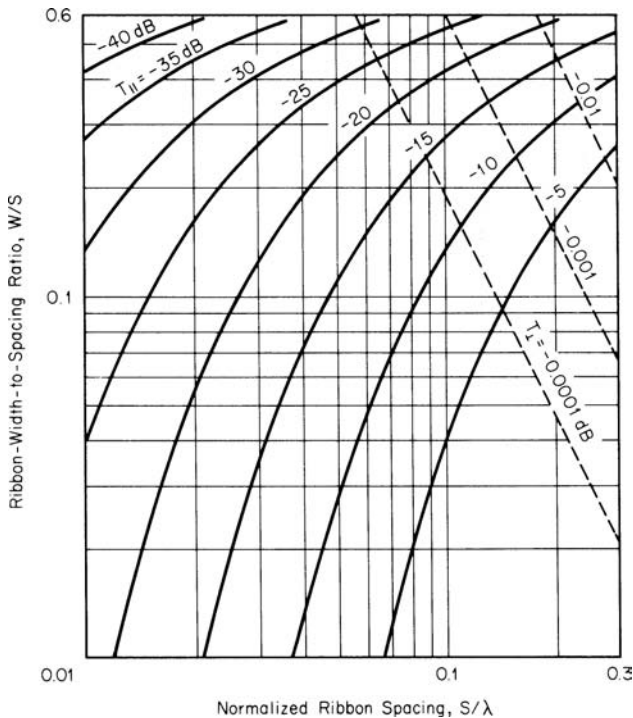


FIGURE 55-2 Transmission coefficient for parallel polarization and for perpendicular polarization for a grid of broadside ribbons of width W and center-to-center spacing S

Increased attenuation of the transmitted signal is obtained by orienting the strips edgewise in the direction of propagation. The length of parallel-plate waveguide below cutoff is greater, resulting in greater attenuation than for the round wires or broadside strips. Figure 55-3 gives the percentage of transmitted power for infinitely long edgewise strips of width W , center-to-center spacing S , and depth t .⁸

The polarization sensitivity of the preceding wire and strip gratings can be eliminated by using (1) a grid of wires composed of two crossed gratings, preferably contacting each other or at least closely spaced, (2) a metal sheet periodically perforated by round or square holes,⁹⁻¹⁰ or (3) expanded metal mesh¹¹ (see Figure 55-4).

The previous structures are relatively broadband devices. Narrowband structures can be constructed by using a periodic array of strip-metal dipoles.¹² Operation below the resonant frequency of the dipoles results in little transmission loss versus essentially total reflection at resonance. Dipole and crossed-dipole arrays are used in dichroic subreflectors for dual-frequency operation for a reflector antenna. Frequency-selective surfaces may also be used to scan the far-field beam of an antenna.¹³

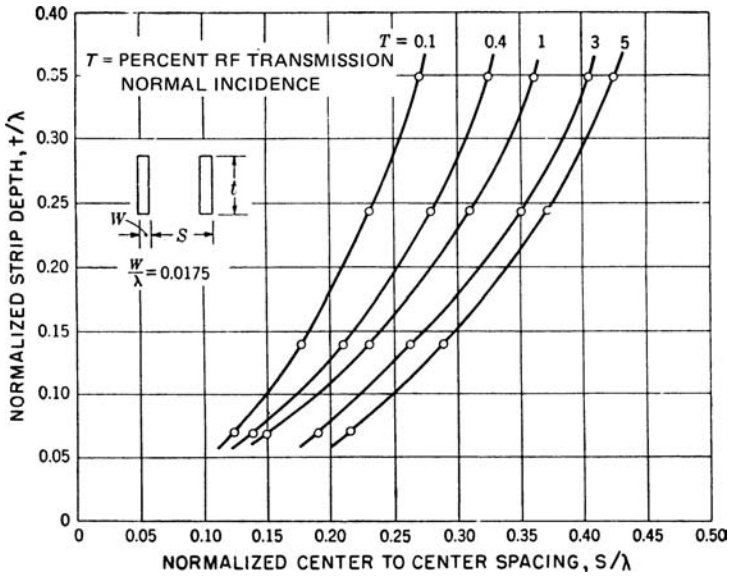


FIGURE 55-3 Grating of edgewise strips: relation between strip depth and spacing for constant transmission⁸

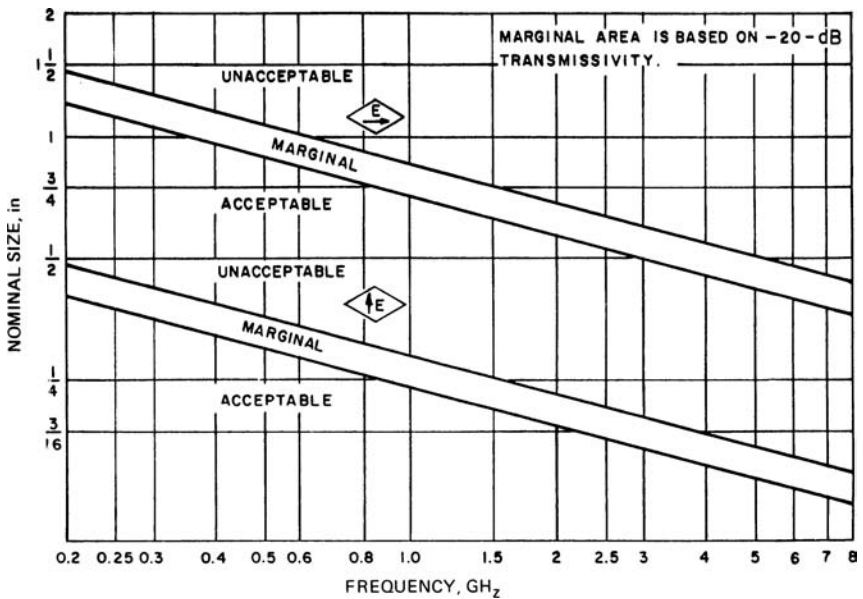


FIGURE 55-4 Radio-frequency properties of expanded-metal-mesh screens¹¹

Transmission-Type Circular Polarizers¹⁴

Transmission-type polarizers are anisotropic media which, when placed over a radiating aperture, convert linearly polarized signals to circularly polarized signals. They may take several forms, including those in the following list:

- Parallel-plate structures using metal slats¹⁵
- Alternating dielectric plates of differing materials¹⁶
- Lattice structures using strips or rods¹⁷⁻¹⁸
- Meander-line photoetched structures¹⁹
- Etched-dipole arrays

These types of polarizers have been applied to linearly polarized antennas when other means of achieving circular polarization cannot be economically or easily achieved. For example, simple horns, biconical horns, as well as slot, horn, and dipole arrays have been successfully equipped with these devices. Recent advances in precision-photoetching technology permit close-tolerance photoetched structures, which exhibit excellent uniformity of insertion phase and loss characteristics. This technology has also been designed for use over greater-than-3:1-frequency bandwidths and with power-handling densities of up to 100 W/in² (average). Structures 1, 2, and 3 are older approaches; the reader is referred to the References for background design information.

Structures 4 and 5, mentioned previously, consist of bonded sandwiches of alternating etched sheets of Mylar or Teflon-fiberglass copper-clad dielectric material; for example, MIL-P-631 Type G Mylar or MIL-P-13949 Type GT Teflon-fiberglass. (See Figure 55-5 for more information.) Bonding is done by using a low-loss bonding film (such as 3 M Type

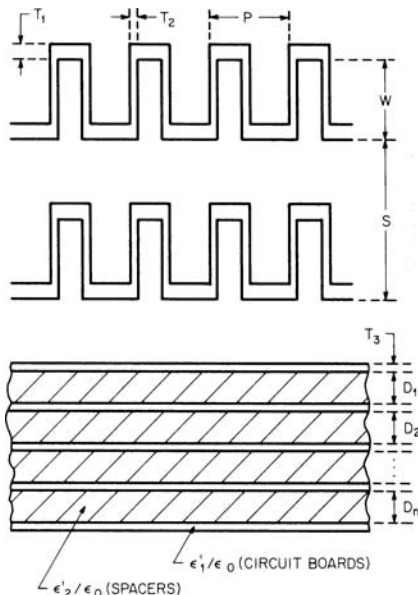


FIGURE 55-5 Geometry of a meander-line circular polarizer showing meander-line and sandwich dimensions

HX-1000) and spacer material such as Hexcell Type HRH-10 honeycomb or Eccofoam SH polyurethane foam. A dielectric layer is generally added to these sandwiches for outer-surface environmental protection; this may consist of epoxy-fiberglass Type G-100 approximately 0.010 to 0.015 in (0.254 to 0.381 mm) thick. Structures of this type have been used on both flat and cylindrical aperture with essentially identical performance.

In use, these structures are oriented so that the incident electric field is 45° to the slow axis of the polarizer. That component of the field, which lies in the slow or perpendicular direction, travels through an equivalent circuit which consists of a ladder network of shunt capacitive elements. The parallel-field component travels through a network of shunt inductive elements. In meander-line polarizers, which are in predominant use today, the shunt capacitance elements arise between the parallel meandering strips²⁰ and, for some very broadband designs, also from added dipole elements which can extend the frequency bandwidth to higher frequencies for a given meander line. Shunt inductive elements are realized by meandering copper lines etched from the copper-clad material. Because of the broadband nature of the multiplayer meander-line polarizer, tolerances on the etching, alignment, and bonding of the sandwich are somewhat looser than for earlier devices such as the metal-slat or dielectric polarizers.

Etched multilayer polarizers are usually designed to be optimum in plane-wave field conditions so that they can achieve better than 2-dB axial ratios, insertion losses of less than 0.2 dB, and a return loss of better than 17 dB over more than octave bandwidths. When they are applied to apertures less than approximately one wavelength in size or when they are forced into the flares of small horns, performance can be expected to deteriorate considerably. In addition, as with all circular polarizers, the best axial-ratio performance will be achieved when provisions are made for dissipating the cross-polarized wave which scatters from the polarizers and reenters the antenna. The latter can be accomplished by the use of a terminated orthomode coupler or by the use of a resistive sheet inside the feeding horn and aligned with the H plane.

Reflection-type polarizers can be derived from the full-transmission-type polarizers described previously, by using half the number of layers, which are then mounted onto a metal sheet.¹³ For narrower-band applications, simple etched-construction transmission-type polarizers can be designed by using narrow-line-width strips as in Lerner²¹ or two-layer arrays of etched dipoles. In the latter, phase-shift quadrature is achieved by the shunt capacitance of a dipole medium only; the differential phase shift due to the longitudinal component of the incident field is negligible. For two layers, axial ratios of less than 0.5 dB can be achieved for bandwidths of a few percent.

Multilayer polarizers are best designed by formulating a model of the susceptance of the metallic circuits on each layer²²⁻²⁵ and then iterating the parameters of a ladder network by using an ABCD-matrix (or equivalent) analysis on a computer. Analysis is performed for both parallel and perpendicular field components. The network's insertion phase and reflection properties are then optimized for a minimum reflection coefficient for both field components and for an insertion phase-shift differential as close as possible to the ideal 90° . Such an analysis can readily model the thickness of the sandwich materials, their electrical properties, the incidence angle, and the frequency-varying nature of the individual susceptances. Reference 21 is useful in this respect. Examples of three documented broadband polarizers are tabulated in Table 55-4 and Figure 55-5.

Wind Loading

The primary forces to which an antenna is subjected are wind, ice, vibration, and acceleration. The wind-loading force is given in Figure 55-6 for solid and open surfaces.²⁸ Ice loading increases the stress on a structure because of the increase in areas presented to wind and because of the additional weight due to the ice loading [approximately 57 lb/ft^3 (913 kg/m^3)].

TABLE 55-4 Parameters of Meander-Line Polarizers Shown in Figure 55-5

	Polarizer 1 ²²	Polarizer 2 ²⁴	Polarizer 3 ²⁵
Frequency range	4.0–8.0 GHz	7.0–14.0 GHz	2.6–7.8 GHz
Axial ratio (maximum)	3.0 dB	3.0 dB	3.0 dB
Return loss increase (polarizer)	5.0 dB	3.0 dB	3.0 dB
Power handling (Average W/in ²)	100	...	50
No. of circuits	4	5	6
Circuit board			
Material	MIL-P-631 Mylar
Thickness (T3)	0.010	0.003
Dielectric constant	2.52	3.00
Spacers			
Material	Foam	Styrofoam	HRH-10-OX Honeycomb
Thicknesses (D_n)	0.200, 0.300, 0.200	All of 0.194	All 0.312
Dielectric constant	1.08	1.08
Outer circuit			
S	0.707	0.570	0.750
E	0.280	0.154	0.215
P	0.3535	0.118	0.058
	0.018	0.031	0.007
T2	0.018	0.017	0.007
Next circuit			
S	Same as outer circuit	Same as outer circuit with dipoles* 0.063 by 0.500 and spaced 1.500 added on alternate meander lines
W		
P		
T1		
T2		
Inner circuit			
S	0.070	Same as outer circuit	Same as outer circuit with dipoles 0.063 by 0.500 and spaced 1.500 added on each meander line
W	0.370		
P	0.3535		
T1	0.027		
T2	0.027		
	0.015 G-10 fiberglass window attached

*Dipoles oriented in the w direction
Note: All dimensions are in inches.

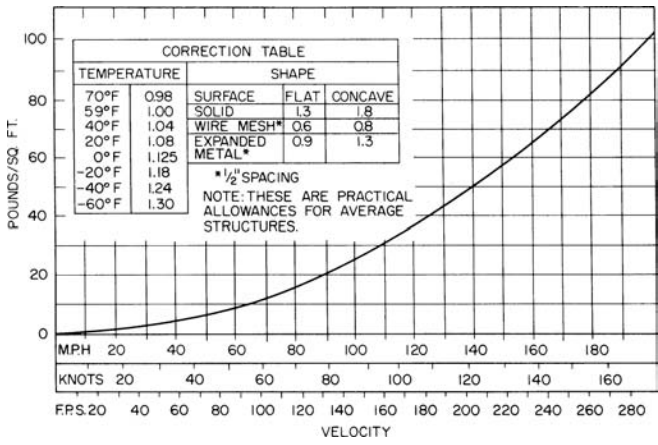


FIGURE 55-6 Wind-loading data

55.4 RADIO-FREQUENCY-BAND DESIGNATIONS

Historically, letter band designations originated during World War II for secrecy reasons. These designations were continued after they had been made public, but different authors used different definitions for the band edges. One set of common-usage band designations is listed in Table 55-5. Standardized bands have been defined by two official international organizations and are also listed in Table 55-5. The Institute of Electrical and Electronics Engineers (IEEE) standard is the preferred radar band designation. The International Telecommunications Union (ITU) assigns frequency allocations by international treaty. Specific frequency ranges within the ITU bands shown in Table 55-5 are assigned for radar use, and these ranges differ slightly depending on which region of the world is involved. ITU band designations are used as a general indication of the frequency of operation.

In addition, the U.S. Department of Defense has issued a directive concerning the designation of frequencies used in electronic-countermeasures (ECM) operations. Each of the 13 letter bands listed in Table 55-5 is further subdivided into 10 equal-bandwidth channels and given a phonetic and number designation such as Delta-4 (1.3 to 1.4 GHz). It is improper usage to designate a radar by an ECM band designation since the ECM designations only apply to ECM equipment.

55.5 ANTENNA-PATTERN CHARACTERISTICS

Both the initial design of an antenna and the validation of measured antenna data are greatly facilitated through the use of theoretically derived pattern characteristics. The patterns produced by a variety of one- and two-dimensional aperture distributions have appeared in the literature over the years. The task of engineers is to determine which theoretical aperture distribution best fits their physical situations so that they can assess antenna performance from the theoretical data. This can be done by measuring or calculating the actual aperture distribution and then comparing it with a variety of theoretical distributions.²⁹⁻³⁰ The three primary factors of concern to antenna designers are gain, beamwidth, and sidelobe level.

TABLE 55-5 Frequency-Band Designations

IEEE Radar Bands*		ITU Frequency Bands†		Common-Usage Bands‡		Electric-Countermeasures Bands§	
Band	Frequency Range, GHz	Band	Frequency Range, GHz	Band	Frequency Range, GHz	Band	Frequency Range, GHz
HF	0.003–0.03	HF	0.003–0.03	HF	0.003–0.03	A	0–0.25
VHF	0.03–0.3	VHF	0.03–0.3	VHF	0.03–0.3	B	0.25–0.5
UHF	0.3–1	UHF	0.3–3	UHF	0.3–1	C	0.5–1
L	1–2	SHF	3–30	L	1–2	D	1–2
S	2–4	EHF	30–300	S	2–4	E	2–3
C	4–8			C	4–8	F	3–4
X	8–12			X	8–12.4	G	4–6
Ku	12–18			Ku	12.4–18	H	6–8
K	18–27			K	18–26.5	I	8–10
Ka	27–40			Ka	26.5–40	J	10–20
mm	40–300			Q	33–50	K	20–40
				V	50–75	L	10–60
				W	75–110	M	60–100

*From Institute of Electrical and Electronic Engineers Standard (Nov. 30, 1976): 521.

†From International Telecommunications Union, Art. 2, Sec. 11 (Geneva, 1959).

‡No official international standing.

§From AFR 55-44 (AR 105–86, OPNAVIST 3430.9B, MCO 3430.1) (October 1964).

This section presents data on these three parameters for both line-source and circularly symmetric distributions.

The gain of an antenna is usually compared with that of an antenna with the same aperture dimensions but having a constant phase and amplitude distribution (i.e., uniformly illuminated). The power gain G of the antenna then is

$$G = G_0 \eta_a$$

where

$G_0 = 4\pi A/\lambda^2$ = gain of uniformly illuminated aperture

A = antenna aperture area

λ = wavelength

$\eta_a = \eta\eta_L$ = antenna efficiency

η = aperture illumination efficiency

η_L = all other efficiency factors

The aperture illumination efficiency represents the loss in gain resulting from tapering the aperture distribution in order to produce sidelobes lower than those achievable from a uniform illumination. For a circular aperture η is given by a single number. For separable rectangular aperture distributions it equals the product (sum in decibels) of the efficiencies in each of the two aperture directions.

The half-power beamwidth BW of an antenna is related to the beamwidth constant β by⁶

$$BW = 2 \sin^{-1} \left(\frac{\beta \lambda}{2L} \right) \approx \beta \frac{\lambda}{L} \quad (\text{linear aperture})$$

$$BW = 2 \sin^{-1} \left(\frac{\beta \lambda}{2D} \right) \approx \beta \frac{\lambda}{D} \quad (\text{circular aperture})$$

where L is the length of the linear aperture and D is the diameter of the circular aperture. The small-argument approximation for the arcsine is typically used for calculating BW. Values of β and η will be given as a function of sidelobe level for a number of distributions in the following subsections.

Continuous Line-Source Distributions

The problem of determining the optimum pattern from a line source has received considerable attention. The optimum pattern is defined as one that produces the narrowest beamwidth measured between the first null on each side of the main beam with no sidelobes higher than the stipulated level. Dolph³¹ solved this problem for a linear array of discrete elements by using Chebyshev polynomials. If the number of elements becomes infinite while element spacing approaches zero, the Dolph pattern becomes the optimum continuous line-source pattern³² given by

$$E(u) = \cos \sqrt{u^2 - A^2}$$

where

$$u = \frac{\pi L}{\lambda} \sin \theta$$

λ = wavelength

θ = angle from the normal to aperture

The beamwidth constant β in degrees and the parameter A are given by

$$\beta = \frac{360}{\pi^2} \sqrt{[\operatorname{arccosh}(R)]^2 - \left[\operatorname{arccosh} \left(\frac{R}{\sqrt{2}} \right) \right]^2}$$

where

$$A = \operatorname{arccosh}(R)$$

R = main-lobe-to-sidelobe voltage ratio

This Chebyshev pattern provides a useful basis for comparison even though it is physically unrealizable since the remote sidelobes do not decay in amplitude. In fact, all sidelobes have the same amplitude in the Chebyshev pattern. The aperture distribution which produces the Chebyshev pattern has an impulse at both ends of the aperture³³ and produces very low aperture efficiency. In addition, the pattern is very sensitive to errors in the levels of these impulses.

Taylor³² developed a method for avoiding the previous problems by approximating the Chebyshev pattern arbitrarily closely with a physically realizable pattern. Taylor approximated the Chebyshev uniform sidelobe pattern close to the main beam but let the wide-angle sidelobe decay in amplitude. Taylor used a closeness-parameter \bar{n} in his analysis. As \bar{n} becomes infinite, the Taylor distribution approaches the Chebyshev distribution. By using the largest \bar{n} that still produces a monotonic aperture distribution, one obtains the beamwidth constant and aperture efficiency shown in Figure 55-7 and Figure 55-8. Notice that the beamwidth from this Taylor distribution is almost as narrow as that from the Chebyshev distribution while still producing excellent aperture efficiency.

Several other common distributions^{8,34} are also listed in Figures 55-7 and 55-8. The advantage of the $\cos^n(\pi x/L)$ distribution and the $\sin(\sqrt{u^2 - B^2})/\sqrt{u^2 - B^2}$ pattern is that both the distribution and the pattern for them may be obtained in closed form. This mathematical convenience is obtained at the expense of poorer beamwidth and efficiency as compared with the Taylor distribution.

Continuous Circular-Aperture Distributions

The Chebyshev pattern of the preceding section can also be shown to be optimum for the circular aperture. Taylor has generalized his line-source distribution to the circular case,³⁵⁻³⁶ and his pattern approaches the Chebyshev pattern as his closeness-parameter \bar{n} for the circular aperture approaches infinity. The beamwidth constant and aperture efficiency shown in Figures 55-9 and 55-10 are obtained by using the largest \bar{n} that still produces a monotonic distribution.³⁷

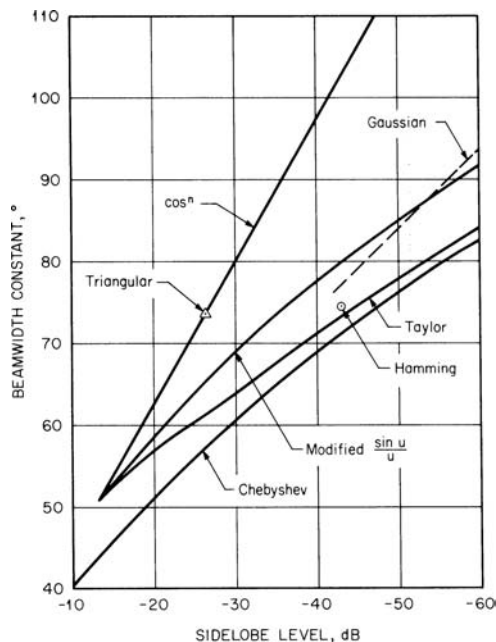


FIGURE 55-7 Beamwidth constant versus sidelobe level for several line-source aperture distributions

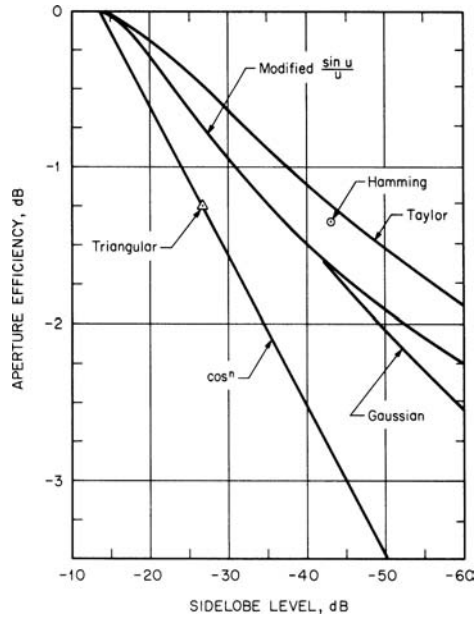


FIGURE 55-8 Aperture efficiency versus sidelobe level for several line-source aperture distributions

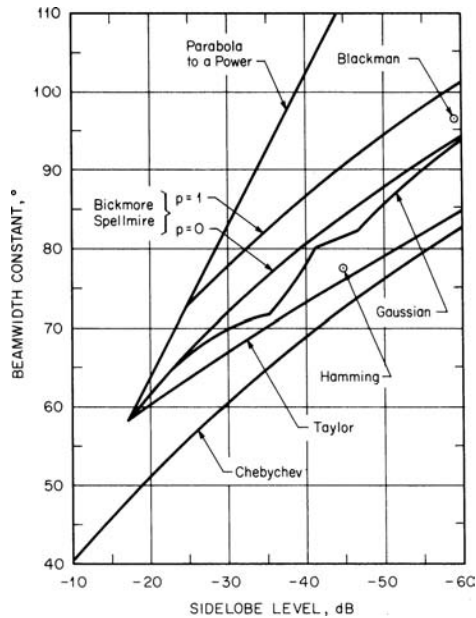


FIGURE 55-9 Beamwidth constant versus sidelobe level for several circular-aperture distributions (after A. C. Ludwig³⁷ © GRC 1981)

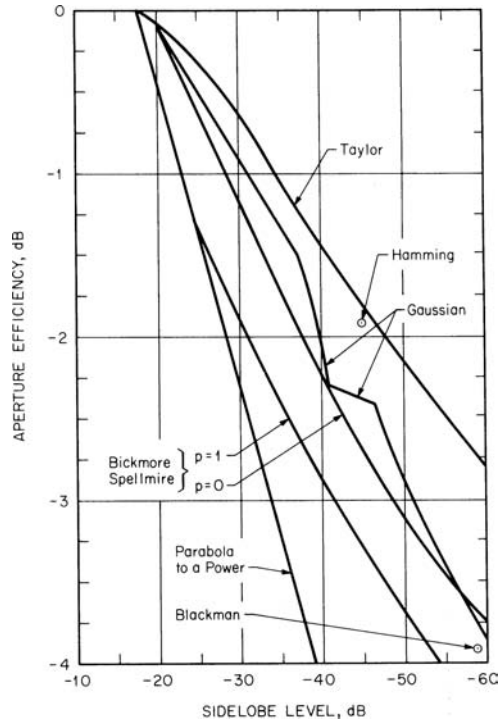


FIGURE 55-10 Aperture efficiency versus sidelobe level for several circular-aperture distributions (after A. C. Ludwig³⁷ © GRC 1981)

The Bickmore-Spellmire distribution³⁸ is a two-parameter distribution which can be considered a generalization of the parabola to a power distribution.³³ The Bickmore-Spellmire distribution $f(r)$ and pattern $E(u)$ are given by

$$f(r) = p \left[1 - \left(\frac{2r}{D} \right)^2 \right]^{p-1} \Lambda_{p-1} \left(jA \sqrt{1 - \left(\frac{2r}{D} \right)^2} \right)$$

$$E(u) = \Lambda_p \left(\sqrt{u^2 - A^2} \right)$$

where p and A are constants that determine the distribution and Λ is the lambda function

$$u = \frac{\pi D}{\lambda} \sin \theta$$

The Bickmore-Spellmire distribution reduces to the parabola to a power distribution when $A = 0$ and to the Chebyshev pattern when $p = -1/2$.

A Gaussian distribution³⁷ produces a no-sidelobe Gaussian pattern only as the edge illumination approaches zero. In general, the aperture distribution must be numerically

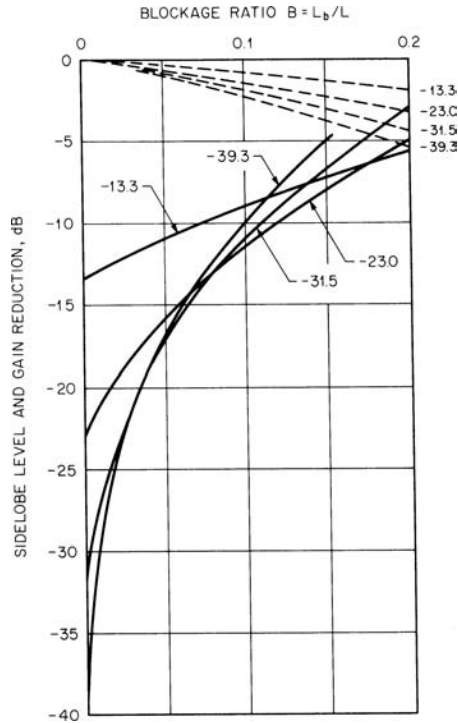


FIGURE 55-11 Gain loss (---) and resulting side lobe level (—) for a centrally blocked line-source distribution having a specified unblocked side-lobe level

integrated to obtain the far-field pattern. The second sidelobe of this pattern is sometimes higher than the first, which accounts for the erratic behavior of β and η in Figures 55-9 and 55-10.

Blockage

The placement of a feed in front of a reflector results in blockage of part of the aperture energy. In the geometric-optics approximation, no energy exists where the aperture is blocked, and the undisturbed aperture distribution persists outside the blocked region.³⁹ Using a line source of length L with a \cos^n aperture distribution and a centrally located blockage of length L_b produces the gain loss and the resulting side lobe level shown in Figure 55-11. The corresponding changes are shown in Figure 55-12 for a circular aperture of diameter D and a parabola to a power distribution blocked by a centrally located disk of diameter D_b . Notice that the line-source blockage affects the pattern much more rapidly than in the circular case since it affects a larger portion of the aperture. Calculation of strut-blockage effects is given by Gray.⁴⁰

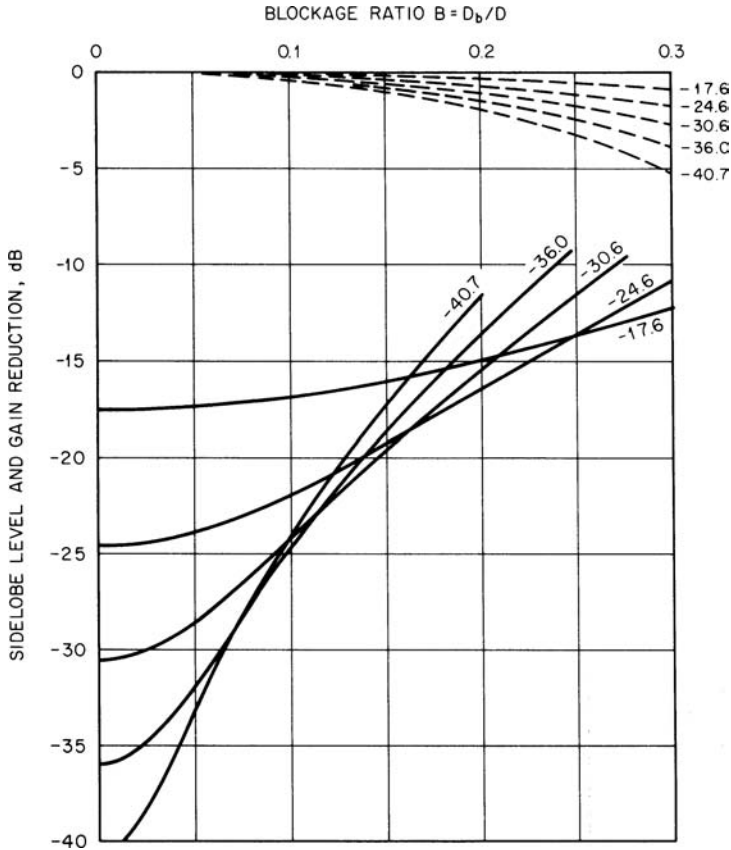


FIGURE 55-12 Gain loss (---) and resulting sidelobe level (—) for a centrally blocked circular-aperture distribution having a specified unblocked sidelobe level

55.6 ELECTRICALLY SMALL HORNS

Aperture theory does not accurately predict the pattern characteristics of electrically small horns (on the order of a wavelength or less in size). Diffraction effects from the flange or rim around the horn (also considered to be edge currents flowing on the outside surface of the horn) markedly influence the pattern. Experimental beamwidth data⁸ for small horns without a flange are given in Figure 55-13. These curves were obtained from measurements of a larger number of 10-dB beamwidths. Empirical formulas for these *E*- and *H*-plane curves are given, respectively, by

$$BW_E(10 \text{ dB}) = 88^\circ \frac{\lambda}{B} \quad \text{for } \frac{B}{\lambda} < 2.5$$

$$BW_H(10 \text{ dB}) = 31^\circ + 79^\circ \frac{\lambda}{A} \quad \text{for } \frac{A}{\lambda} < 3$$

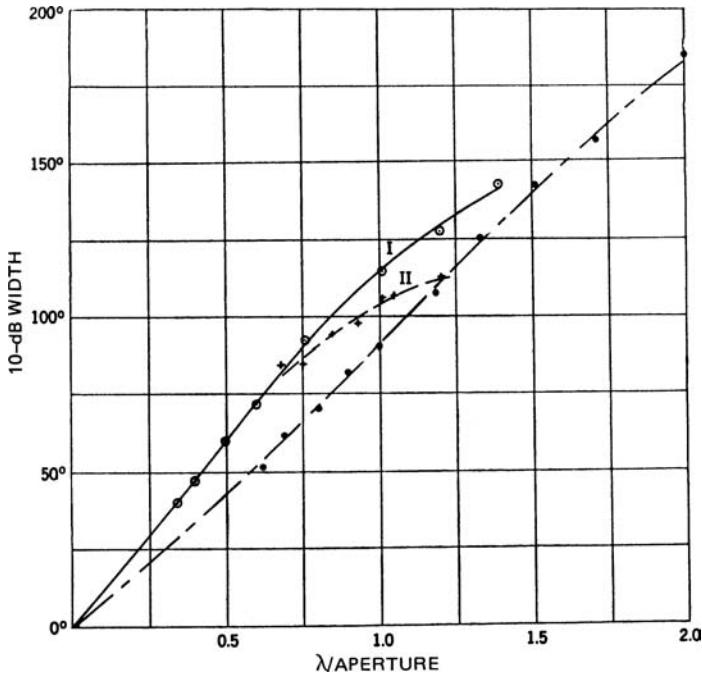


FIGURE 55-13 Experimental 10-dB beamwidths of horns having small phase variations over the aperture ($\frac{\delta}{\lambda g} < \frac{1}{8}$): ———— E plane; - - - - H -plane sectoral horns; ———— H plane of compound horns with E -plane aperture equal to or greater than a wavelength⁸

where B and A are the E - and H -plane aperture dimensions, respectively. The actual beamwidth of any particular horn may vary from the previous values since different flare angles produce different phase variations over the aperture. Moment Method computer codes can be used to accurately predict the pattern of these horns.

REFERENCES

1. J. D. Walton (ed.), *Radome Engineering Handbook* (New York: Marcel Dekker, Inc., 1970).
2. J. Agranoff (ed.), *Modern Plastics Encyclopedia* (New York: McGraw-Hill, 1981).
3. A. R. Von Hippel (ed.), *Dielectric Materials and Applications* (Cambridge, Mass: The M.I.T. Press, 1954).
4. J. J. Gau, W. D. Burnside, and M. Gilreath, "Chebyshev Multilevel Absorber DESIGN concept," *IEEE Trans. Antennas Propagation*, vol. 45 (August 1997): 1286–1293.
5. D. G. Bodnar, "SEASAT Antenna Study," Final Tech. Rep. EES/GIT, Proj. A-1617-000, Georgia Institute of Technology (August 1974): 27–32.
6. T. Larsen, "A Survey of the Theory of Wire Grids," *IRE Trans. Microwave Theory Tech* (May 1962): 191–201.
7. W. W. Mumford, "Some Technical Aspects of Microwave Hazards," *IRE Proc* (February 1961): 427–445.

8. S. Silver (ed.), *Microwave Antenna Theory and Design*, M.I.T. Rad. Lab. Ser., vol. 12 (New York: McGraw-Hill, 1949).
9. W. D. Hayes, "Gratings and Screens as Microwave Reflectors," MIT Radiation Laboratory, Report (April 1943): 54–20.
10. P. W. Hannan and P. L. Burgmyer, "Metal-Grid Spatial Filter," Interim Tech. Rep. RADC-TR-79-295, AD No. A089756, Hazeltine Corporation (July 1980).
11. *ITE Antenna Handbook*, 2nd Ed. (Philadelphia: ITE Circuit Breaker Company, 1965): 26.
12. Chao-Chun Chen, "Scattering by a Two-Dimensional Periodic Array of Conducting Plates," *IEEE Trans. Antennas Propagation*., vol. AP-18 (September 1970): 660–665.
13. F. S. Johnson, "A New Planar Grating-Reflector Antenna," *IEEE Trans. Antennas Propagation*., vol. AP-38 (September 1990): 1491–1495.
14. This subsection was written by John M. Seavey, Adams-Russell Company, Amesbury, Mass.
15. S. Cornbleet, *Microwave Optics* (London: Academic Press, Inc., 1976) 309–323.
16. H. S. Kirschbaum and L. Chen, "A Method of Producing Broad-Band Circular Polarization Employing an Anisotropic Dielectric," *IRE Trans. Microwave Theory Tech.* (July 1957): 199–203.
17. J. A. Brown, "Design of Metallic Delay Dielectrics," *IEE Proc.* part III, vol. 97, no. 45. (January 1950): 45.
18. N. Amitay and A. A. M. Saleh, "Broad-Band Wide-Angle Quasi-Optical Polarization Rotators," *IEEE Trans. Antennas Propagation*., vol. AP-31 (January 1983): 73–76.
19. L. Young, L. A. Robinson, and C. A. Hacking, "Meanderline Polarizer," *IEEE Trans. Antennas Propagation*., vol. AP-23 (May 1973): 376–378.
20. N. Marcuvitz, *Waveguide Handbook* (New York: McGraw-Hill, 1951).
21. D. S. Lerner, "Wave Polarization Converter for Circular Polarization," *IEEE Trans. Antennas Propagation*., vol. AP-13 (January 1965): 3–7.
22. T. L. Blakney, J. R. Burnett, and S. B. Cohn, "A Design Method for Meanderline Circular Polarizers," USAF Antenna Conf., Allerton Park, IL, 1972.
23. A. C. Ludwig, M. D. Miller, and G. A. Wideman, "Design of Meanderline Polarizer Integrated with a Radome," *IEEE Antennas Propagation Int. Symp. Dig.* (1977): 17–20.
24. R. S. Chu and K. M. Lee, "Analytical Model of a Multilayered Meander-Line Polarizer Plate with Normal and Oblique Plane-Wave Incidence," *IEEE Trans. Antennas Propagation*., vol. AP-35 (June 1987): 652–661.
25. C. Terret, J. R. Levrel, and K. Mahdjoubi, "Susceptance Computation of a Meander-Line Polarizer Layer," *IEEE Trans. Antennas Propagation*., vol. AP-32 (September 1984): 1007–1011.
26. D. A. McNamara, "An Octave Bandwidth Meanderline Polarizer Consisting of Five Identical Sheets," *IEEE Antennas Propagation Int. Symp. Dig.* (1981): 237–240.
27. Data furnished courtesy of H. Schlegel, Adams-Russell Company, Amesbury, Mass.
28. H. Jasik (ed.), *Antenna Engineering Handbook*, 1st Ed., Chap. 35 (New York: McGraw-Hill, 1961).
29. R. C. Hansen, *Microwave Scanning Antennas*, vol. 1, Chap. 1 (New York: Academic Press, Inc., 1964).
30. J. F. Ramsay, "Lambda Functions Describe Antenna/Diffraction Pattern," *Microwaves* (June 1967): 69–107.
31. C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Side-Lobe Level," *IRE Proc.*, vol. 34 (June 1946): 335–348.
32. T. T. Taylor, "Design of Line-Source Antennas for Narrow Beamwidth and Low Side Lobes," *IRE Trans. Antennas Propagation*., vol. AP-3 (January 1955): 16–28.
33. J. W. Sherman III, "Aperture-Antenna Analysis," *Radar Handbook*, M. I. Skolnik (ed.), Chap. 9 (New York: McGraw-Hill, 1970).
34. F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *IEEE Proc.*, vol. 66 (January 1978): 51–83.

35. T. T. Taylor, "Design of Circular Apertures for Narrow Beamwidth and Low Sidelobes," *IRE Trans. Antennas Propagat.*, vol. AP-8 (January 1960): 17–22.
36. R. C. Hansen, "Tables of Taylor Distributions for Circular Aperture Antennas," *IRE Trans. Antenna Propagat.*, vol. AP-8 (January 1960): 23–26.
37. A. C. Ludwig, "Low Sidelobe Aperture Distributions for Blocked and Unblocked Circular Apertures," RM 2367, General Research Corporation, April 1981.
38. W. D. White, "Circular Aperture Distribution Functions," *IEEE Trans. Antennas Propagat.*, vol. AP-25 (September 1977): 714–716.
39. P. W. Hannan, "Microwave Antennas Derived from the Cassegrain Telescope," *IRE Trans. Antennas Propagat.*, vol AP-9 (March 1961): 140–153.
40. C. L. Gray, "Estimating the Effect of Feed Support Member Blockage on Antenna Gain and Side-Lobe Level," *Microwave J.* (March 1964): 88–91.

Chapter 56

Frequency Selective Surfaces

Stephen W. Schneider
John F. McCann

Air Force Research Laboratory

CONTENTS

56.1	INTRODUCTION	56-2
56.2	THE PLANE WAVE EXPANSIONTECHNIQUE	56-3
56.3	ANATOMY OF A FREQUENCY SELECTIVE SURFACE: GRIDS AND ELEMENTS	56-8
56.4	STABILITY OF FSS WITH ANGLE OF INCIDENCE AND POLARIZATION	56-17
56.5	OTHER APPLICATIONS	56-21

56.1 INTRODUCTION

The electromagnetic properties of periodic surfaces have been widely studied for some time now. A periodic surface is generally defined as a uniform assembly of elements (or scatterers) arranged in one- or two-dimensional infinite arrays. Several arrays and material layers may be combined to produce resonant structures commonly referred to as a Frequency Selective Surface (FSS). These surfaces find a wide variety of applications over a vast range of the electromagnetic spectrum, from below UHF to the far-infrared regions. In the microwave region, periodic surfaces have been used with great success as phased array antennas,¹⁻³ artificial dielectrics,⁴⁻⁷ diffraction gratings,⁸⁻⁹ frequency selective reflectors for antennas,¹⁰⁻¹¹ dichroic surfaces,¹²⁻¹⁵ angular filters,¹⁶⁻¹⁸ and spatial filters.¹⁹⁻²² While many of the principles and design approaches described in this chapter have direct application to active structures such as phased arrays, the emphasis here is on passive frequency selective surfaces. The purpose of this chapter is to provide an overview of the field rather than a comprehensive technical exposition and therefore the interested reader is encouraged to consult the references cited in this chapter, in particular the review articles²³⁻²⁴ and most importantly, the textbooks.²⁵⁻²⁸

A frequency selective surface is essentially a filter designed to exhibit different reflection and/or transmission properties as a function of frequency. Unlike an ordinary filter, which generally considers only the frequency response, an FSS must maintain bandfilter characteristics for various angles of incidence and for various polarizations. The latter issue is readily addressed by using elements that respond to the particular polarization of interest. The former, however, represents a formidable challenge because, independent of element type, variations in angle of incidence result in both a shift in resonant frequency as well as a change in functional bandwidth (including magnitude and phase response) with incidence angle and polarization.

An FSS generally consists of a periodic arrangement of either wire or slot elements. An FSS of wire elements is composed of arbitrarily shaped conducting elements, which support electric currents. Alternatively, a slot FSS is composed of arbitrarily shaped perforations in a metallic screen, which support magnetic currents. Surfaces comprised of wire elements act as bandstop filters (Figure 56-1a), because the elements are reflective at resonance and they pass waves above and below the dipole resonant frequency. For a given incidence angle and polarization,

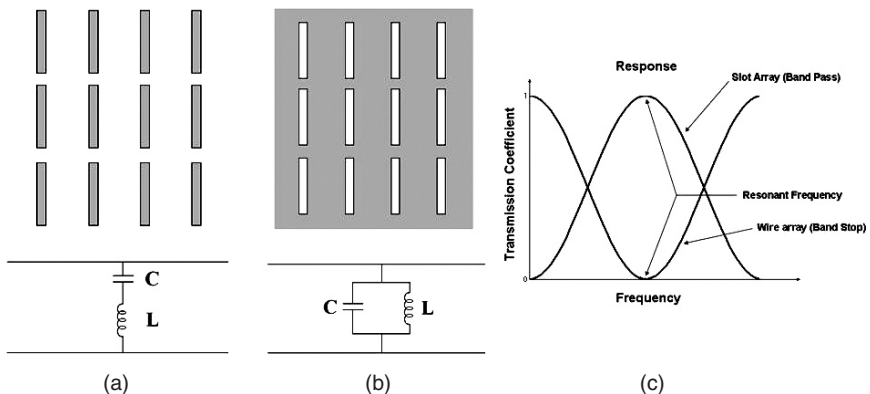


FIGURE 56-1 Generic dipole array (a) and complementary slot array (b) with equivalent transmission line circuits. Typical transmission coefficient curves are shown for both types of surfaces in (c).

an equivalent circuit for a wire FSS consists of a series-tuned LC pair on a transmission line. At resonance the LC combination results in a short circuit (reflects the energy), while away from resonance the LC combination is an open circuit (passes the energy).

Surfaces comprised of slots act as bandpass filters (Figure 56-1b), because the slot elements are transparent at resonance and they reflect waves above and below the slot resonant frequency. For a given incidence angle and polarization, its equivalent circuit is a parallel-tuned LC pair on a transmission line. Here, at resonance, the inductor and capacitor pair combine in order to yield an open circuit response (pass the energy), while away from resonance the LC combination is a short circuit. It is important to reemphasize that while an equivalent LC-circuit may be defined that closely models the behavior of an FSS, the values for the LC-circuit will generally vary greatly with angle of incidence and polarization.

It is convenient to view an FSS as consisting of two distinct regions: a central portion and the portion near the edges. When considering an FSS that consists of a relatively few number of elements, the frequency response is dominated by contributions from the edge of the surface. This means that the influence of the changes in the mutual coupling between all of the elements may easily be found by direct application of conventional computational electromagnetic approaches (Method of Moments,¹³ Finite Element Method,¹⁴ and so on). There has been significant work on finite FSS that addresses the edge effects and the reader is encouraged to consult the resulting publications.²⁸⁻³¹ More typically, Frequency Selective Surfaces are electrically very large (in that they contain many thousands of elements) and the central region of the FSS dominates the response. In this case the vast majority of the elements of an FSS appear to be in the same environment (same impedances and coupling), and it is convenient to analyze the FSS as if it were infinite. Under this condition, certain theoretical approaches may be used that allow insight into the basic behavior of the FSS in a relatively direct manner. The Plane Wave Expansion Technique³² readily lends itself to the analysis of FSS providing a deeper understanding of the nature of the FSS. This deeper understanding results in better design procedures and practices.

While manufacturing defects can lead to suboptimal performance in practical applications,³³ the discussion here assumes infinite arrays under ideal conditions to introduce the reader to the basic approaches and concepts in FSS development. The general outline for this chapter was developed with that purpose in mind. Section 56.2 describes a commonly used analysis approach for FSS—the Plane Wave Expansion Technique. This spectral domain technique provides the fundamental theoretical foundation used in the development of the subsequent material. Section 56.3 describes typical grids and elements that comprise the anatomy of an FSS. To avoid grating lobes, which can severely degrade FSS performance, approaches are presented for realizing compact elements within tight grid spacing. Section 56.4 gives approaches for addressing traditional problems that arise in FSS design. Specifically, dielectric loading techniques and cascading of multiple surfaces are shown to provide stability with angle of incidence and polarization, as well as with bandwidth control and resonance shaping in an FSS. Finally, Section 56.5 provides other applications of FSS, beyond those described previously, and highlights recent trends in this vast field.

56.2 THE PLANE WAVE EXPANSION TECHNIQUE

The plane wave expansion technique is often used to analyze frequency selective surfaces. This approach provides great insight into the scattering problem by applying the theory of antenna arrays in which elements first receive incident energy and then reradiate it to form the transmitted and reflected beams. This section will provide a brief overview of the theory behind this technique.

Consider a two-dimensional (located in the xz -plane) infinite array of linear elements with orientation \hat{p} , and interelement spacings D_x and D_z , as shown in Figure 56-2a. Now suppose the array is exposed to an incident plane wave propagating in the direction:

$$\hat{s} = \hat{x}s_x + \hat{y}s_y + \hat{z}s_z \quad (56-1)$$

The direction components are defined as follows:

$$\begin{aligned} s_x &= \sin \eta \cos \alpha \\ s_z &= \sin \eta \sin \alpha \\ s_y &= \sqrt{1 - s_x^2 - s_z^2} \end{aligned} \quad (56-2)$$

In Eq. 56-2, η is the elevation angle with respect to the y -axis and α is the azimuth angle with respect to the x -axis, as shown in Figure 56-2b.

One can assume without loss of generality that a reference element centered at the origin $(0,0,0)$ has a current distribution defined as $I_{00}(l)$ induced by the plane wave. For an infinite array, Floquet's theorem requires that the element in the q th column and m th row of the array has a current distribution, $I_{qm}(l)$, of the same amplitude as the reference element, but with a linear phase variation resulting from the incident plane wave impinging on the displaced (relative to the origin) element. This element current is given by

$$I_{qm}(l) = I_{00}(l) e^{-j\beta_q D_x s_x} e^{-j\beta_m D_z s_z} \quad (56-3)$$

where values $\beta_q = 2\pi q/\lambda$ and $\beta_m = 2\pi m/\lambda$ are the phase velocities of the incident plane wave along the array.

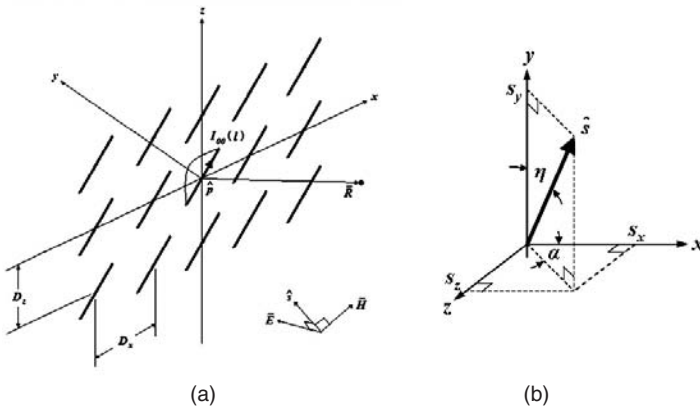


FIGURE 56-2 (a) Infinite planar array of linear elements with orientation \hat{p} and interelement spacings D_x and D_z . Also shown is a reference element at the origin with current distribution $I_{00}(l)$, and an observation point \vec{R} . (b) The incident field direction \hat{s} as defined by the angles α and η .

The electric field resulting from these current excitations at an observation point \bar{R} is

$$\bar{E}(\bar{R}) = I_{00}(0,0) \frac{Z}{2D_x D_z} \sum_k \sum_{\eta=-\infty}^{\infty} \frac{e^{-j\beta\bar{R} \cdot \hat{r}_{\pm}}}{r_y} \bar{e}_{\pm} P \quad (56-4)$$

where Z is the intrinsic impedance of the medium surrounding the array. In this equation, (k,n) denote the spectral indices rather than the element location (q,m) . The pattern function, P , is derived from the current distribution of the element as follows:

$$P = \frac{1}{I_{00}(0,0)} \int_{\text{element}} I_{00}(l) e^{j\beta\hat{p} \cdot \hat{r}_{\pm}} dl \quad (56-5)$$

The field vector, \bar{e}_{\pm} , is defined as

$$\bar{e}_{\pm} = [\hat{p} \times \hat{r}_{\pm}] \times \hat{r}_{\pm} \quad (56-6)$$

Finally, the direction vector \hat{r}_{\pm} is defined as

$$\hat{r}_{\pm} = \hat{x} \left(s_x + k \frac{\lambda}{D_x} \right) \pm \hat{y} s_y + \hat{z} \left(s_z + n \frac{\lambda}{D_z} \right) \quad (56-7)$$

with

$$r_y = \sqrt{1 - \left(s_x + k \frac{\lambda}{D_x} \right)^2 - \left(s_z + n \frac{\lambda}{D_z} \right)^2} \quad (56-8)$$

where λ is the wavelength in the surrounding medium. The derivation from Eq. 56-3 to Eq. 56-4 is rather complicated, and involves a transformation using Poisson's sum formula to change the summation variables from the element (spatial) indices q and m to the spectral indices k and n . The details of this derivation are omitted here, but may be found elsewhere.²⁷

Upon examination of the term $e^{-j\beta\bar{R} \cdot \hat{r}_{\pm}}$ in Eq. 56-4, which describes a plane wave emanating from the origin, it is evident that only real values of r_y will result in propagating waves. (r_x and r_z will always be real.) Imaginary values of r_y will result in what are called evanescent waves, which attenuate exponentially as the observation point moves away from the array. When $(k,n) = (0,0)$, there are propagating waves in the direction of the incident field (both forward and backward scattering for positive and negative values of r_y). There can also be propagating waves for values of $(k,n) \neq (0,0)$; these are grating lobes and are determined by both the interelement spacing in terms of wavelengths and the direction of the incident plane wave. The evanescent waves present throughout the spectral domain represent stored energy close to the array, which greatly impacts the array impedance as well as the transmission and reflection properties of the periodic surface.³⁴

It follows from Eq. 56-4 that the voltage induced on a test element located at \bar{R} by the currents of the infinite periodic array is

$$V = I_{00}(0,0) \frac{Z}{2D_x D_z} \sum_k \sum_{\eta=-\infty}^{\infty} \frac{e^{-j\beta\bar{R} \cdot \hat{r}_{\pm}}}{r_y} \hat{p}' \cdot \bar{e}_{\pm} P' P \quad (56-9)$$

where \hat{p}^i is the orientation of the test element and P^i is the pattern function of the test element. From there, the mutual impedance between the array and the test element is derived as

$$Z_M = \frac{Z}{2D_x D_z} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{e^{-j\beta \hat{R} \cdot \hat{r}_z}}{r_y} \left[{}_{\perp} P {}_{\perp} P^i + {}_{\parallel} P {}_{\parallel} P^i \right] \quad (56-10)$$

where the pattern functions have been decomposed into parallel (\parallel) and perpendicular (\perp) components. Now if the test element is located a wire radius a from the array reference element (in the y -direction only, such that $\hat{R} = \hat{y}a$), the array self-impedance is found to be

$$Z_A = \frac{Z}{2D_x D_z} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{e^{-j\beta a r_y}}{r_y} \left[{}_{\perp} P {}_{\perp} P^i + {}_{\parallel} P {}_{\parallel} P^i \right] \quad (56-11)$$

If there are no grating lobes, this array self-impedance can be broken down into real and imaginary components rather simply. Under this condition, it was shown before that the only propagating waves will occur when $(k,n) = (0,0)$; this is the only term in the spectral domain with a real component, which can be denoted as R_A . The rest of the terms of Z_A (i.e., $(k,n) \neq (0,0)$) will result in purely imaginary contributions and can be summed along with the imaginary component of the $k = n = 0$ term to obtain the reactance X_A . Therefore the array impedance can be expressed as

$$Z_A = R_A + jX_A \quad (56-12)$$

If only the principal planes are considered, with no grating lobes, it can be shown that wire arrays of simple elements can be modeled by an equivalent transmission line, as shown in Figure 56-3.²⁷

From Figure 56-3, it is deduced that the reflection coefficient looking into the array can be expressed as

$$\Gamma = -\frac{R_A}{R_A + jX_A} = -\frac{1}{1 + \frac{jX_A}{R_A}} \quad (56-13)$$

While the transmission line model and reflection coefficient expression have been shown for a simplified case, these prove to be very useful tools in the design of any frequency selective surface comprised of wire arrays. The basic concepts shown here can be expanded to surfaces of complex elements, surfaces that employ multiple FSS, and surfaces that employ dielectric layers. For instance, if instead there are multiple arrays in the

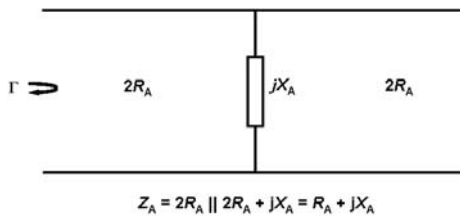


FIGURE 56-3 Equivalent circuit of wire array, valid for no grating lobes and principal planes only

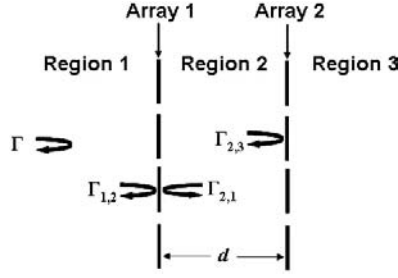


FIGURE 56-4 Double-array structure. The arrays are separated by a distance d , and the three regions have the same material properties.

same structure, such as the double array case shown in Figure 56-4, the scattering properties can easily be determined from well-known analysis of stratified media. The reflection coefficient looking right into the structure in Figure 56-4 is found to be²⁷

$$\Gamma = \frac{\Gamma_{1,2} + \Gamma_{2,3} e^{-j2\beta d} + (\Gamma_{1,2} + \Gamma_{2,1})\Gamma_{2,3} e^{-j2\beta d}}{1 - \Gamma_{2,1}\Gamma_{2,3} e^{-j2\beta d}} \quad (56-14)$$

where the three regions are comprised of the same medium (β is the same), d is the distance between the arrays, and the reflection coefficients of the arrays are defined as shown in Figure 56-4.

At first glance, Figure 56-4 makes it appear that $\Gamma_{1,2}$ and $\Gamma_{2,1}$ would be equal and opposite and therefore would cancel each other out in the third term of the numerator of Eq. 56-14. This would be the case if the reflection coefficients are associated with a dielectric interface, or that of widely separated wire arrays; however, in general terms this should not be assumed. If there are two or more arrays within the near field regions of each other, or if slot arrays are being analyzed (more on that later), $\Gamma_{1,2} \neq -\Gamma_{2,1}$. This is the result of interaction between the arrays. The nearfield region is defined as within the point at which the first evanescent plane wave has decayed to 1% of its original value at the array surface. This is determined to be³⁴

$$\frac{-\ln(0.01)}{\beta \sqrt{(\lambda/D_{\max})^2 - 1}} \quad (56-15)$$

where D_{\max} is the maximum of the interelement spacings. If the arrays under question are spaced outside of this nearfield region, only the propagating modes need be considered. However, when the arrays are located within the nearfield region, evanescent modes must be included in the calculation of the array mutual coupling. This mutual coupling is simply found through an extension of Eq. 56-10, and is given by³⁴

$$Z_{N,M} = \frac{-Z_0}{2D_x D_z} \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{e^{-j\beta(\bar{R}^N - \bar{R}^M) \cdot \hat{r}_z}}{r_y} \left[\hat{p}^N \cdot \bar{e}_z^M \right] P^N P^M \quad (56-16)$$

This is the mutual impedance of array M on array N , centered at \bar{R}^M and \bar{R}^N , with element orientations \hat{p}^M and \hat{p}^N , and having element pattern functions P^M and P^N , respectively.

The intrinsic impedance of the surrounding medium is given as Z_0 , D_x and D_z are the interelement spacings of array N , and the field vector \vec{e}_\pm^M is defined similarly to Eq. 56-6:

$$\vec{e}_\pm^M = [\hat{p}^M \times \hat{r}_\pm] \times \hat{r}_\pm \quad (56-17)$$

This expression for the mutual impedance is used along with the array self-impedances (see Eq. 56-11) to construct the full impedance matrix in order to find currents along the array elements. From there, it is a simple step to calculate the scattered electric field. Examples of surfaces with multiple arrays will be explored later in this chapter.

56.3 ANATOMY OF A FREQUENCY SELECTIVE SURFACE: GRIDS AND ELEMENTS

An FSS is formed by the repetition of identical structures in two dimensions. Therefore an FSS is comprised of two components: a grid (often referred to as a lattice) which specifies the periodicity of the surface, and a reference element associated with every grid point. The grid/element combination, used in concert with tuning materials, define the frequency, bandwidth, polarization response, angular coverage, power handling capabilities, cost and mechanical durability of the FSS. As such it is important to understand the basic functionality and limitations of some common configurations used in FSS design, recognizing that obtaining good FSS performance is not assured by optimizing only performance parameters. In this section, commonly used reference elements and grids are described.

Grating Lobes

A fundamental problem with all FSS is that grating lobes will emerge when the frequency/angle of incidence combination becomes high enough for a given set of interelement spacings, D_x and D_z . To illustrate this, consider a plane wave incident on an array of elements, as shown in Figure 56-5. As described in the previous section, the incident plane wave induces currents on the elements of the FSS. In an infinite array, the current magnitude is the same for each element, while the phase differs by a linear progression (Floquet's theorem). These currents then reradiate to form the scattered fields in the forward direction (transmission) and backward direction (reflection). When the elements are closely spaced, the phase will not vary greatly from element to element and hence the contribution from each element will coherently add together in only the specular direction. These fields are given by the $(k,n) = (0,0)$ term in Eq. 56-4. Because the phase is cyclic, if the separation between the elements is large, the reradiated currents can add in other ambiguous directions in addition to the specular directions. These ambiguous directions are known as grating lobes. Depending on the resonant frequency and pattern factor of the array elements, the grating lobe direction may contain greater power density than the specular direction.

The onset of grating lobes is determined by inspection of Eqs. 56-4 and 56-8, indicating that grating lobes exist at angles represented by (s_x, s_z) given that

$$(s_x + k\lambda/D_x)^2 + (s_z + n\lambda/D_z)^2 = 1 \quad (56-18)$$

In Eq. 56-18, circles are defined with unit radius and centers $(k\lambda/D_x, n\lambda/D_z)$ in the $s_x - s_z$ plane. As shown in Figure 56-6 for $D_x = 0.75\lambda$ and $D_z = 0.49\lambda$, when these circles

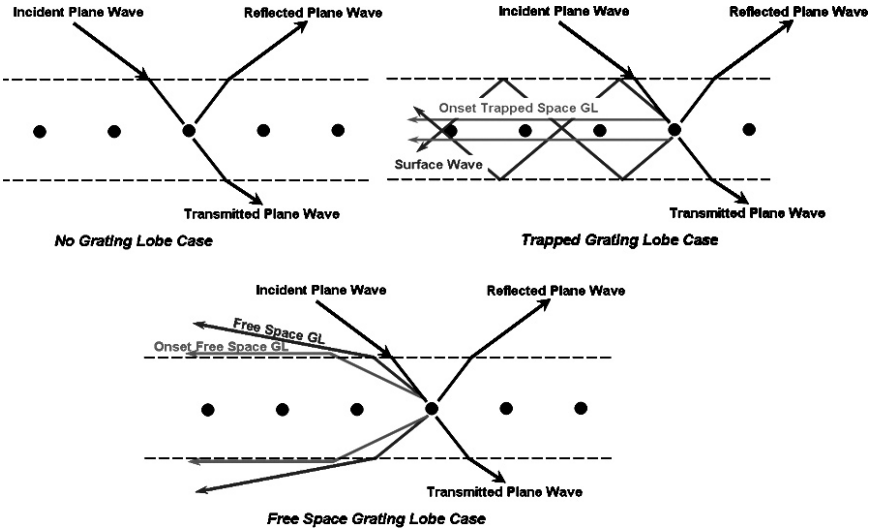


FIGURE 56-5 As the interelement separation distance in an FSS gets larger, grating lobes will exist.

overlap, grating lobes for the corresponding values of s_x and s_z are obtained.³⁵ Note that the onset of grating lobes for a given angle of incidence depends only on the frequency (λ) and the interelement spacing (D_x and D_z). Further, the emergence of grating lobes in free space is independent of the presence of a dielectric media. When the FSS is inside a dielectric substrate, the electrical spacing between the elements is larger so the condition for grating lobes occurs precisely as if the entire space was filled with the dielectric (that is, at a lower frequency). As a consequence of Snell's law, total internal reflection of these grating lobes occurs until the actual spacing of the elements exceeds the onset of free space grating lobes defined in Eq. 56-18. These internal grating lobes are referred to as *trapped*

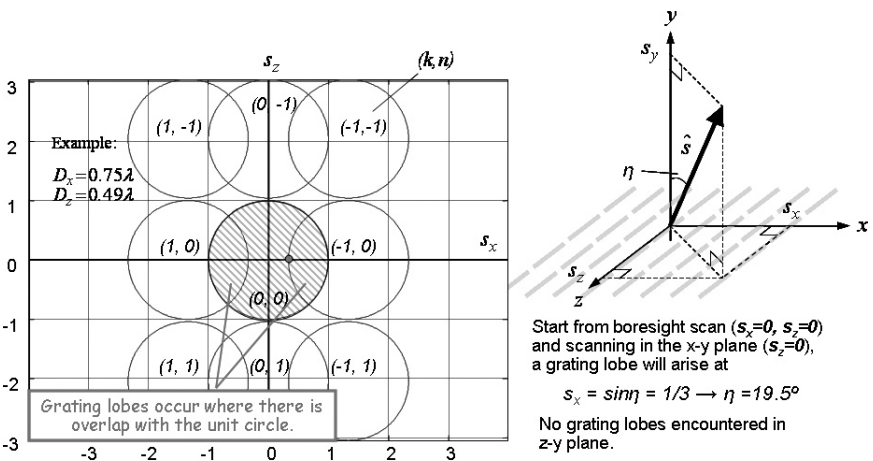


FIGURE 56-6 A graphic representation of Eq. 56-14 yields a grating lobe diagram.

grating lobes and are shown in Figure 56-5. Most of these trapped grating lobes have little impact on the FSS performance, because they are finite and do not propagate in the substrate unless they are driven (these cases are referred to as “forced” surface waves). In between the onset of trapped grating lobes and free space grating lobes, the onset of free surface waves may occur. These surface waves are dependent on the incidence angles and geometry of the dielectric substrates, and are essentially trapped grating lobes guided by the dielectric slab. This results in a lack of energy propagation to any other propagating modes. Like the onset of free space grating lobes, free surface waves are characterized by anti-resonances (nulls) in the reflection curves, which have a devastating impact on the FSS performance and should, in general, be avoided. Significant work has been done on surface wave phenomena and their impact on FSS and more details may be found in the resulting publications.^{27,31,36–37}

In principle, avoiding grating lobes is simply a matter of making the interelement spacing sufficiently small (ideally less than one-half wavelength in the dielectric). To do this effectively requires a basic understanding of the surface grid and elements and the ways that they impact the performance of the FSS.

Grids

The grid (or lattice) plays a significant role in the frequency response of an FSS. The orientation and density with which the elements of an FSS are configured impacts the bandwidth, resonant frequency, polarization response, as well as the onset of grating lobes. Symmetry is an important consideration in FSS design, in particular when low cross-polarization is required.³⁴ FSS geometries possessing greater than a two-fold rotational symmetry have reflection (transmission) coefficients that are independent of polarization at normal angle of incidence.³⁸ Furthermore, no cross-polarization component can exist for any angle of incidence in a plane of symmetry. While the elements in an FSS may have any degree of rotational symmetry (for example, a circular loop element), the grid cannot. A consequence of the crystallographic restriction theorem is that two dimensional grids (lattices) can possess only one-, two-, three-, four-, or six-fold rotational symmetry,³⁹ because the grid is required to be both rotationally and translationally invariant.

There are four grid types commonly used in a planar FSS: rectangular, regular skewed, square, and equilaterally triangular. Examples of these grid types are shown in Figure 56-7. From this figure one can observe that the rectangular and regular skewed grids have two-fold rotational symmetry, the square grid has four-fold rotational symmetry, and the equilaterally triangular grid has six-fold rotational symmetry. The planes of symmetry are indicated in

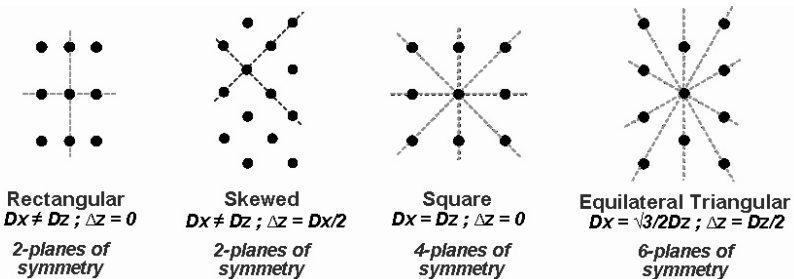


FIGURE 56-7 Four types of grids, possessing various degrees of symmetry, are typically used in FSS.

the figure. The overall degree of symmetry that an FSS possesses will depend on the grid/element combination; hence the choice of grid for a given application is highly dependent on the reference element used. Specifically, the grid/element combination should provide a lattice spacing small enough to avoid grating lobes in the frequency band of interest, while providing the maximum amount of symmetry.

In addition to delaying the onset of grating lobes, reducing the grid spacing can greatly increase the operational bandwidth of an FSS. To illustrate this, consider the widely studied “gangbuster” surface^{41–42} formed by rotating straight wire (dipole) elements in a square grid until the ends of the reference element point toward a grid location in an adjacent column ($x + D_x$) and offset an integer number of rows ($z + nD_z$), as shown in Figure 56-8. Since the elements are in a square grid, as n increases the interelement separation is smaller for the same length of dipole (as indicated in the figure insert). The reflection coefficient at normal angle of incidence for type-2, type-3, and type-4 gangbuster surfaces is shown in Figure 56-8 for a fixed reference element size. The results indicate that as the elements are more densely packed, the resultant bandwidth is increased. It is further noted that the resonant response of the surface remains quite constant, differing by less than 10% in all three cases for the same length dipole.

A careful examination of the array self-impedance term in Eq. 56-11 provides the rationale for the bandwidth increase. Note in this equation that the double summation is multiplied by a $1/D_x D_z$ term. As D_x and D_z are made smaller, the real part of the array self impedance, R_A , gets larger while the imaginary part X_A remains about the same.²⁷ The result is that $R_A \gg X_A$ and the reflection coefficient of Eq. 56-13 goes to $\Gamma = -1$ over a broader range of frequencies.

Elements

The reference element is a primary determinant of the resonant frequency, bandwidth, polarization response, and angular coverage of an FSS. While endless configurations of reference elements may be considered, the choice of element ultimately depends on the intended application of the FSS. Regardless of application, the elements of an FSS should have a stable resonant frequency with angle of incidence and should be electrically compact enough to fit within a grid that avoids both free space and trapped grating lobes. Reducing the size of a reference element generally requires loading techniques. Loading of the element is typically accomplished by encapsulating the element in a material (dielectric loading), reshaping the element (element loading), or some combination of both techniques.

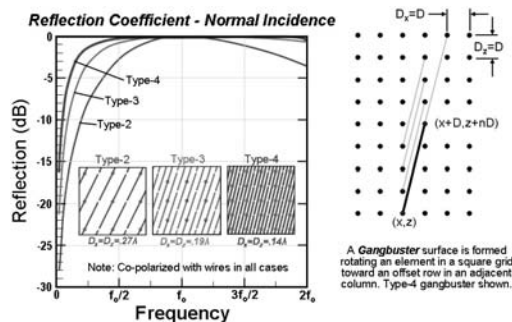


FIGURE 56-8 Gangbuster surfaces of various types are shown to illustrate how reducing the grid spacing increases the operating bandwidth of an FSS.

Antenna engineers have long recognized that immersing an antenna in a dielectric material provides an effective means of reducing the size of the element. It is well known that when an FSS is immersed in an infinite medium of relative dielectric constant ϵ_r , the resonance is scaled to the frequency $f_1 = f_o / \sqrt{\epsilon_r}$, where f_o is the resonant frequency in free space. Therefore, if all dimensions of an FSS inside the dielectric are reduced by a factor of $1/\sqrt{\epsilon_r}$, the resonant frequency will remain the same. When a relatively thin dielectric substrate ($\sim 0.05\lambda_o$) is added to both sides of any FSS, all resonances will remain effectively the same as they would if the dielectric were infinite. This is because the electric field is relatively strong in the immediate vicinity of these resonant elements due to the evanescent waves (see Eq. 56-4). This resonance shift for thin dielectrics substrates is, for all intents and purposes, the same for wire and slot arrays. Furthermore, as discussed previously, the onset of free space grating lobes will occur at a higher frequency (by a factor of $\sqrt{\epsilon_r}$) than without a dielectric substrate (although as previously discussed, care must be taken to avoid free surface waves). In addition to controlling the resonant frequency of an FSS, thicker dielectric slabs ($0.25\lambda_o$ or more) are commonly used to control the shape and variation of the resonant response with angle of incidence and frequency (see Section 56.4). Contrary to the thinner slabs, thicker slabs have a different effect on the behavior of slot arrays as compared to wire arrays.

Another well-known approach for reducing the size of the elements in an FSS is applying reactive loading to the elements themselves. As shown previously in Figure 56-1a, a dipole array is seen to be essentially equivalent to a network of inductive and capacitive elements connected in series. At resonance they produce a short (or act like a ground plane producing a reflection coefficient equal to -1). As the elements are made shorter, the FSS becomes capacitive. To tune the elements back to resonance at a reduced size, reactive terms may be added. Two fundamental approaches are available: inductive center-loading and capacitive end-loading. Figure 56-9 shows the approximate physical length and equivalent circuit for these arrays of loaded dipole elements using each of these techniques. While loading may be accomplished by introducing lumped loads into the FSS, the traditional technique for loading the elements is to change the shape of the elements as will now be described.

As shown in Figure 56-9 (middle), the equivalent circuit for inductive center-loading is basically the same as a reference array of straight dipoles (Figure 56-9, left) with the exception that a larger inductance has been added to the middle of the element (denoted by more coils in the schematic) to compensate for the capacitive array reactance. This distributed inductance may be synthesized by attaching a short-circuited, two-wire, transmission line stub across the terminals of the shortened dipole, as shown in Figure 56-9 (middle). The overall length of a center-loaded dipole will be approximately a third of a wavelength, as opposed to the half wavelength required for the reference straight dipole. The reflection coefficient of an inductive, center-loaded dipole array observed at broadside and 60°

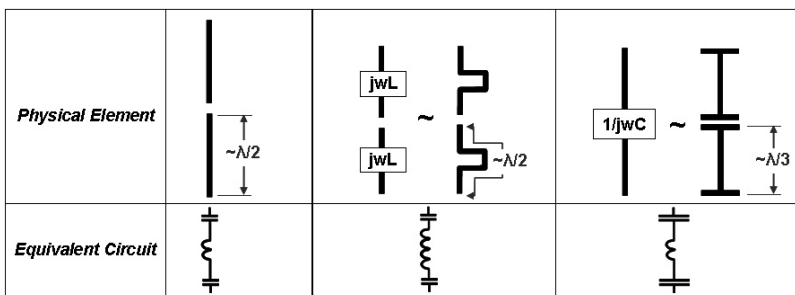


FIGURE 56-9 Loading may be used to reduce the size of a reference element at resonance. (Left) a reference dipole, (center) a center-loaded dipole, (right) an end-loaded dipole.

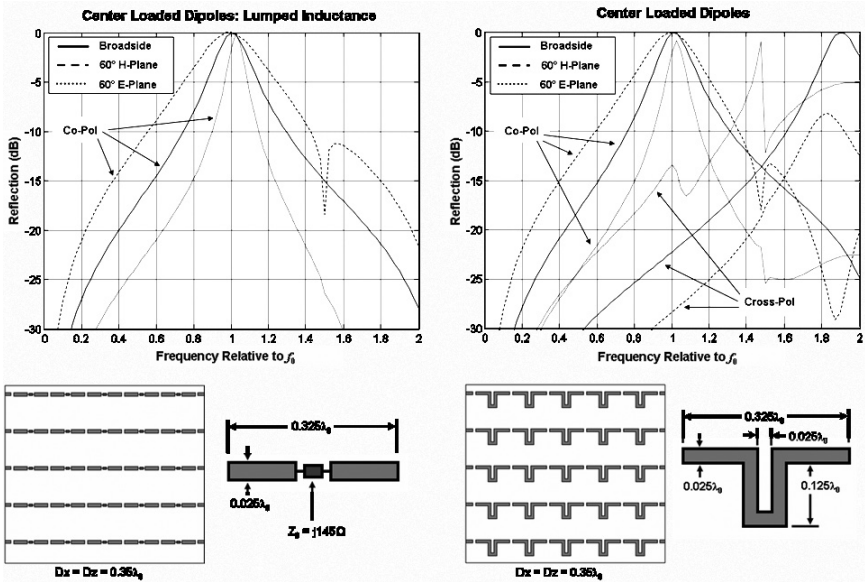


FIGURE 56-10 Comparison of the reflection coefficients for a lumped (left) and distributed (right) center-loaded dipole. Note that these elements resonate at $\sim \lambda/3$ length.

from the surface normal, for both E - and H -planes, as a function of frequency, is shown in Figure 56-10 for both the lumped (left) and distributed cases (right). It is clear from this figure that while the distributed load provides a similar shift in the resonant frequency to the lumped load, the structure associated with the distributed load has a very real contribution to the frequency response of the FSS. Specifically, as the frequency increases, the stub itself radiates, giving rise to the large cross-polarization component observed in Figure 56-10 (right). To eliminate this cross-polarization component, two mirrored center-loaded dipole elements can be connected together to form a four-legged loaded element,²⁷ because the individual cross-polarization components of each element will cancel.

A similar approach may be applied to the end-loaded dipole of Figure 56-9 (right). Here, the equivalent circuit has an inductance approximately equal to the reference array of dipoles; however, the capacitance has been significantly increased as indicated by the longer capacitor symbol. This equivalent circuit is seen to produce a lower resonant frequency (or a smaller element for the same resonance frequency). As shown in the figure, this capacitance may be obtained by the addition of an end-loading that creates a strong capacitive coupling with its neighboring element. This capacitive end-loading results in an element that is approximately a third of a wavelength, in contrast to the reference half-wavelength, straight dipole. Just as in the center-loaded case, the structure associated with the end-loading will contribute to the overall frequency response of the FSS, which should be accounted for in the element design.

In the previous discussion it was shown that reshaping the element may provide the size reduction necessary to produce an FSS operating without grating lobes. It is also clear (from Figure 56-10) that introducing this shaping by distributed loading can significantly impact the response of an FSS. This is one of the primary reasons that the large number of different element types exist for FSS. Examples of typical elements used in FSS design are shown in Figure 56-11. These elements may be broadly grouped into

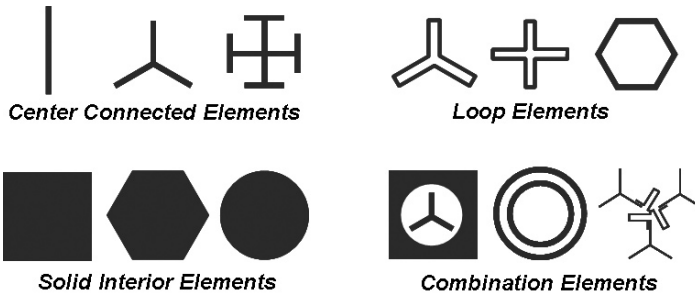


FIGURE 56-11 Typical reference element types arranged in categories (after Munk²⁷ © Wiley 2003). Many combinations are possible.⁴³⁻⁴⁶

four categories²⁷—center-connected elements, loop elements, solid interior elements, and combination elements. Before discussing each of these categories, it is important to note that the electrical properties of FSS can vary dramatically due to the influence of element shape and mutual coupling.

Center-connected elements (or n -pole elements, where n is the number of legs) are formed by attaching radial “legs” from the center of the element. Examples from this category include: dipole, tripole (or three-legged element), anchor elements (or three-legged top-loaded element), and the Jerusalem cross (or four-legged top-loaded element). These elements resonate (for unloaded cases) when their length is $\sim \lambda/4$ from the center connection point to the tip of a radial leg. The simplest example of this class of elements is the straight dipole (shown in Figure 56-12 for a type-4 gangbuster surface). While straight dipole elements are capable of extremely large bandwidths because they may be packed infinitely close together (Figure 56-8), most unloaded n -pole elements are capable of modest bandwidth because they fit into a grid spacing that is just greater than a quarter of a wavelength. Typically, the parallel and perpendicular polarization patterns differ dramatically with these types of elements due to even and odd mode interactions in the parallel plane (a.k.a. modal interaction). This interaction of the even and odd mode also gives rise to a modal interaction null (Figure 56-12 insert). This is a consequence of Foster’s Reactance

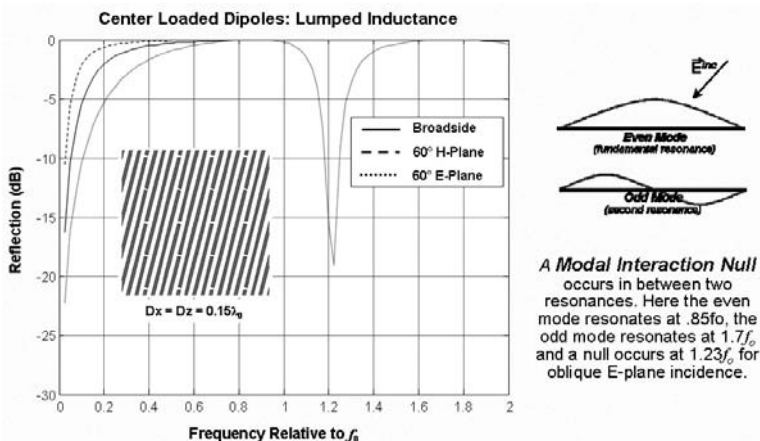


FIGURE 56-12 A type-4 gangbuster surface is an example of a center-connected element. Note the modal interaction null in the E-plane between the fundamental and odd mode resonance.

Theorem,⁴⁷ which states that a null must exist between two zeros (for a lossless system). There is little that can be done with this null (an exception is to add loss to the system), but proper design of the overall FSS performance can allow the null to be placed at a frequency where the impact is small.

Examples from the loop element category include: three- and four-legged loaded elements,²⁷ square loops,⁴⁸ hexagonal loops,³⁴ and circular loops.⁴⁹ Although these elements look quite different, they all behave like loop elements with a fundamental resonance when the circumference is $\sim 1\lambda$. The second resonance occurs when the element has a circumference of $\sim 2\lambda$ for both planes of incidence. Due to their inherently small size, these elements may be packed in a grid spacing $\sim \lambda/3$. This tight grid and the lack of an odd mode second resonance makes them capable of wide band performance. Typically, the parallel and perpendicular polarized patterns of this class of elements differ very little, making them good elements for applications requiring circular polarization. Figure 56-13 shows an FSS of four-legged loaded elements formed by connecting two center-loaded dipoles (see Figure 56-10). Note that a grating lobe occurs for this configuration at $\sim 1.4f_0$ at oblique angles of incidence.

Although among the first types of elements to be investigated, by and large the solid interior elements yield poor performance and are consequently not useful for most FSS applications. Their principle shortfall is that, as most antenna engineers know, fat elements generally do not resonate. This surface can be made to resonate through element interaction with the grating lobe mechanism, but this is typically undesirable.

The final category of element is known as combination elements. These elements are generally designed to exploit one or more of the techniques discussed previously, in order to create a frequency response tailored to an application. For example, it is well known that all elements have natural resonant responses, and that they may not fall where they are needed to satisfy a required band of operation. In between these resonances a null will occur as a result of Foster's Reactance Theorem. Modification of an element by connecting various loads or dissimilar elements in order to shift these resonances leads to a virtually endless number of configurations that can be considered.

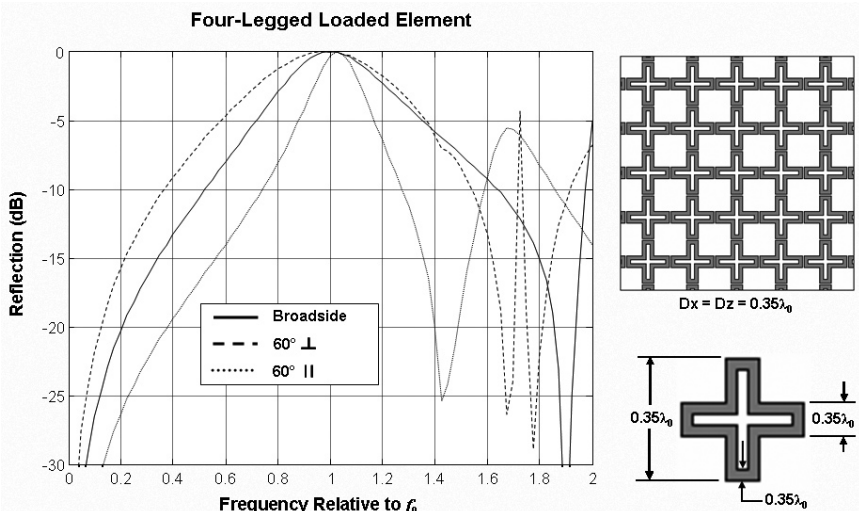


FIGURE 56-13 A four-legged loaded element is an example of a loop element. The loop is made by connecting two inductively center-loaded dipoles.

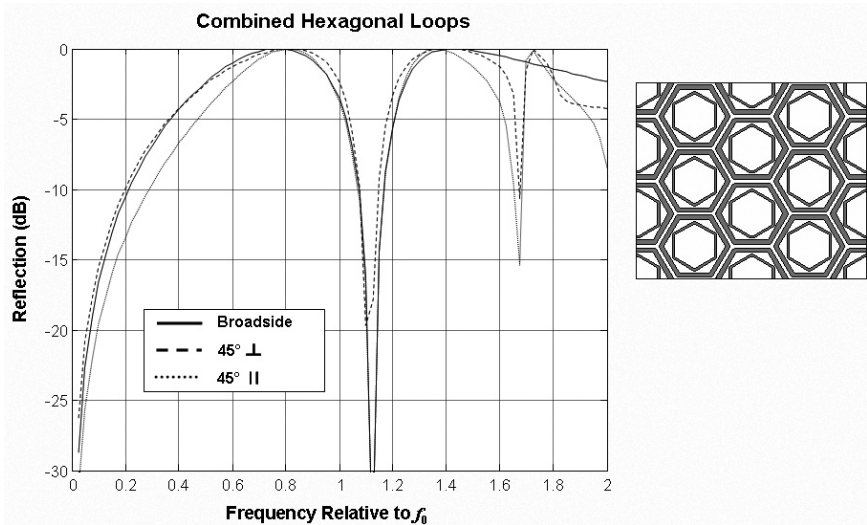


FIGURE 56-14 Concentric hexagonal elements are an example of a combination element. This element produces two resonances with a null in between as a result of Foster's Reactance Theorem.

Figure 56-14 shows an example combination element FSS, consisting of concentric hexagonal loop elements used to produce two resonances at $0.8f_0$ and $1.4f_0$.

Slot Arrays

Slot arrays can be understood as complementary surfaces of wire arrays. A complementary surface has conductor and aperture interchanged relative to the original design (see Figure 56-1). Using duality and the equivalence principle, the plane wave expansion technique described earlier in this chapter can be applied to analyze slot arrays. One would simply replace the impedance Z , with admittance Y , and evaluate the pattern factors associated with the elements using magnetic rather than electric currents. The admittance of slot arrays is related to the impedance of complementary wire arrays by the well-known Babinet's principle.

Using this approach in design will produce slot arrays with similar but dual characteristics as wire arrays (admittance and impedance are interchanged; reflection and transmission coefficients are interchanged), but only if the arrays are in free space. For duality to be applicable when dielectric materials are used with wire arrays, materials with relative magnetic permeability equal to the relative electric permittivity (and vice versa) must be used with slot arrays. This is not practical in general, especially over any range of frequencies and with low loss. Although duality does not perfectly apply, the addition of dielectric material to slot arrays does in general result in similar stabilization properties as it does with wire arrays, specifically in terms of scan angle and polarization. This is demonstrated by the example shown in Section 56.4, and is explained in detail in the reference section.²⁷ Also, because of the equivalence principle, a slot array must be considered as two isolated arrays of magnetic current placed on each side of a PEC ground plane. Therefore, two or more cascaded slot arrays will behave much differently than a corresponding cascade of wire arrays. The isolation between adjacent arrays means that cascaded slot arrays will reradiate only from the outer surfaces, whereas cascaded wire arrays will have forward and backward scattering from each array, leading to interference nulls.

All of this must be taken into consideration when employing slot arrays for designing band-pass filters or active antenna arrays. In the active antenna case, the differences between slot arrays and wire arrays becomes painfully clear when a ground plane is introduced.⁵⁰

56.4 STABILITY OF FSS WITH ANGLE OF INCIDENCE AND POLARIZATION

This section introduces two tools used to tailor the performance of FSS. The first is the use of dielectric slabs to stabilize bandwidth over angle of incidence and polarization. The second is cascading two or more FSS to shape the transmission or reflection curve for a flatter pass/stop band and a sharper roll-off.

Upon examination of Eq. 56-11, it can be seen that for cases exhibiting no grating lobes (i.e., $k = n = 0$ is the only propagating mode), the real part of the array self-impedance can be expressed as

$$R_A = \frac{Z}{2D_x D_z} \frac{1}{r_y} \left[{}_{\perp}P {}_{\perp}P' + {}_{\parallel}P {}_{\parallel}P' \right] \quad (56-19)$$

where the exponential term in Eq. 56-11 has been approximated as unity. It is evident from the $1/r_y$ term in Eq. 56-19 that R_A will vary greatly with scan angle. It is also noted that the pattern factor, a property of the elements themselves, contributes to changes in impedance with polarization as well as scan angle. Here R_A will be examined for two different element types: a simple Hertzian element of length Δl and a square loop with circumference of one-wavelength. Table 56-1 shows the pattern factor of these two elements for both perpendicular and parallel polarizations. The pattern factor expressions for the square loop element are found in the reference section.⁵¹ Figure 56-15 shows R_A plotted for these two cases. Each curve is normalized so that R_A is equal to unity at broadside. Looking at the curves for the Hertzian element, one can see that R_A changes with $1/\cos\eta$ for perpendicular polarization and changes with $\cos\eta$ for parallel polarization. For the case of the square loop element, the R_A curves are different than that of the Hertzian element, but the overall shapes are similar and it results in the same conundrum. In short, there is a divergence with scan angle of the terminal impedance of the two polarizations. The rest of this section will demonstrate how to mitigate this effect and stabilize impedance with respect to polarization and scan angle.

Dielectric Slabs

One common method of stabilizing impedance is the use of dielectric slabs. To understand the utility of dielectric slabs, it is instructive to examine an array in a semi-infinite medium, as shown in Figure 56-16. The array is located in the medium with intrinsic impedance Z_2 a distance d_2 from the dielectric interface.

TABLE 56-1 Pattern Factors of Both Perpendicular and Parallel Polarizations for Both a Hertzian Element and a One-wavelength Square Loop Element

	\perp Polarization (<i>H</i> -Plane)	\parallel Polarization (<i>E</i> -Plane)
Hertzian Element	Δl	$\Delta l \cos \eta$
Square Loop	$\cos[(\pi/4)\sin \eta]$	$[\sin \eta \sin[(\pi/4)\sin \eta] - \cos[(\pi/4)\sin \eta]]/\cos \eta$

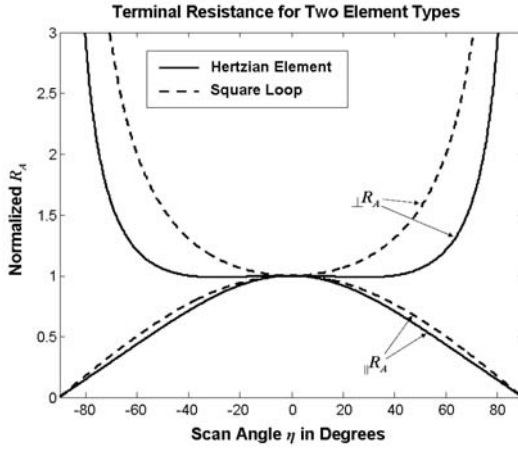


FIGURE 56-15 Terminal resistance R_A plotted against scan angle for both parallel and perpendicular polarizations, for arrays of Hertzian elements and one-wavelength square loop elements.

It can be shown²⁷ that the overall reflection coefficient looking into the dielectric interface can be expressed as

$$\Gamma_{1,2}^a = \frac{\Gamma_{1,2} + \Gamma_A e^{-j2\beta_2 d_2 r_{2y}}}{1 - \Gamma_{2,1} \Gamma_A e^{-j2\beta_2 d_2 r_{2y}}} \quad (56-20)$$

where $\Gamma_{1,2}$ and $\Gamma_{2,1}$ are as shown in Figure 56-16. Γ_A is the reflection coefficient of the array, the simple form of which is given by Eq. 56-13. To simplify Eq. 56-20, it is beneficial to eliminate the exponential terms in the numerator and denominator. To do this, the spacing of the array in the dielectric is chosen such that $d_2 r_{2y} \approx \lambda/4$, leading to $2\beta_2 d_2 r_{2y} \approx \pi$ and

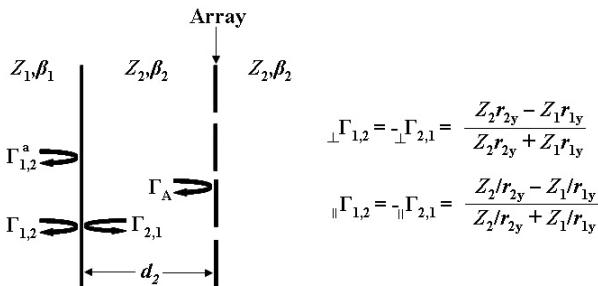


FIGURE 56-16 Array in semi-infinite medium located a distance d_2 from dielectric interface, with reflection coefficient expressions provided

hence $e^{-j\beta_2 d_2 r_{2y}} \approx -1$. Using this approximation, the reflection coefficient of Eq. 56-20 after manipulation can be expressed as

$$\Gamma_{1,2}^a = \frac{1 + \frac{jX_A}{R_A} \frac{\Gamma_{1,2}}{1 + \Gamma_{1,2}}}{1 + \frac{jX_A}{R_A} \frac{1}{1 + \Gamma_{1,2}}} \quad (56-21)$$

If it is assumed that the intrinsic impedances Z_1 and Z_2 are reasonably close in value, then the reflection coefficient $\Gamma_{1,2}$ will be small and the second term in the numerator of Eq. 56-21 can be ignored. Comparison of Eqs. 56-21 and 56-13 shows that the variance of the real part of the terminal impedance, termed the Dielectric Stabilization Factor (DSF),³⁷ for this asymmetric case is

$$_{\perp, \parallel} DSF_A = \frac{1}{_{\perp, \parallel} R_A} \frac{1}{1 + _{\perp, \parallel} \Gamma_{1,2}} \quad (56-22)$$

Figure 56-17 shows the DSF of an array of Hertzian elements (a) and an array of square loop elements (b) for both polarizations as a function of scan angle for different dielectric constants of the second medium. It is assumed that $\epsilon_1 = 1.0$ for each case. The spacing is $d_2 = \lambda/4$ in the dielectric material for each case as well. The case of $\epsilon_2 = \epsilon_1 = 1.0$ corresponds to the array being in free space, and the curves are identical to those shown in Figure 56-15. It is evident that the addition of dielectric material can be utilized to stabilize impedance (and hence reflection and transmission) properties of a FSS. Comparing Figure 56-17a and Figure 56-17b clearly indicates the importance of the pattern factor in determining the stabilization of the response, with respect to the dielectric constant of the medium chosen. The method used here can be extended to find the DSF of the symmetric case (or identical dielectric slabs of both sides of the array).³⁷

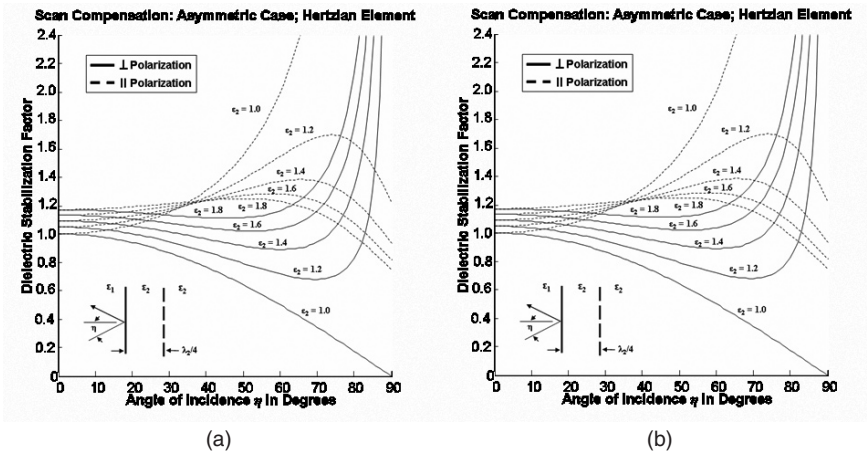


FIGURE 56-17 Asymmetric Dielectric Stabilization Factor (DSF_A) for an array of Hertzian dipoles (a) and an array of square loop elements (b) in a semi-infinite medium placed one quarter-wavelength from dielectric interface

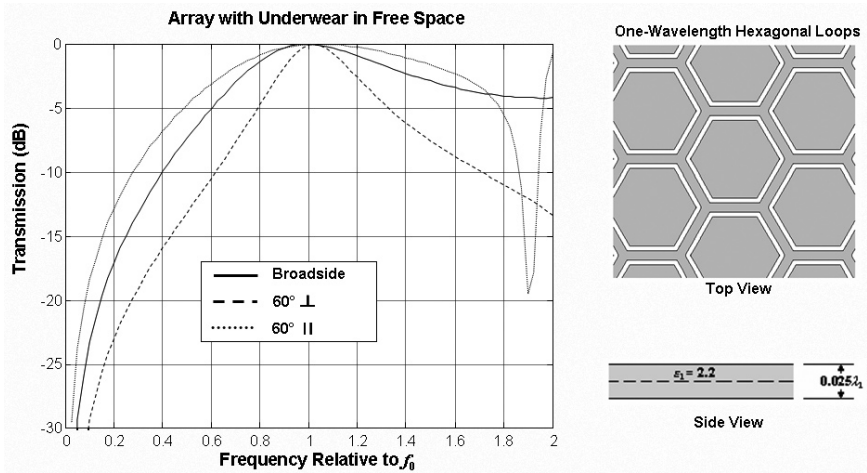


FIGURE 56-18 Transmission curves of FSS of hexagonal slot elements

Figures 56-18 and 56-19 demonstrate this dielectric stabilization. Figure 56-18 shows the transmission properties of an array of hexagonal slot elements. Note the discrepancies of the curves at parallel and perpendicular polarizations at a 60° scan angle. Figure 56-19 shows the transmission properties of the same array but with additional dielectric slabs.

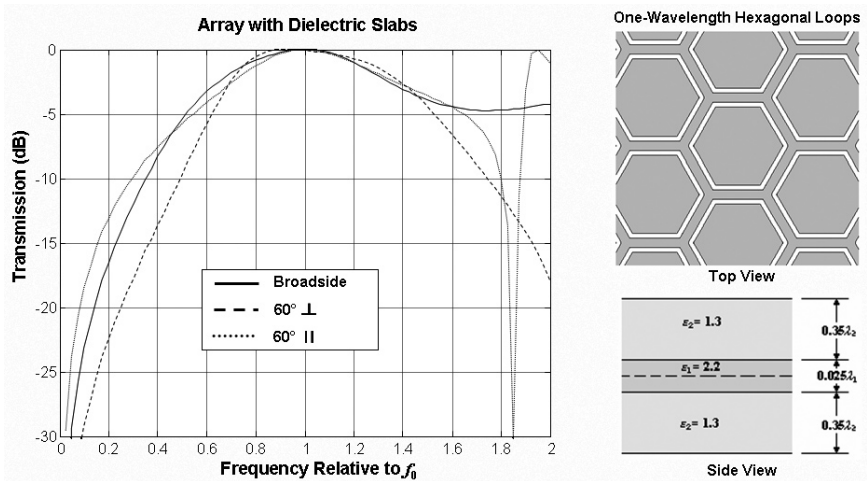


FIGURE 56-19 Transmission curves of FSS of hexagonal slot elements with dielectric slabs

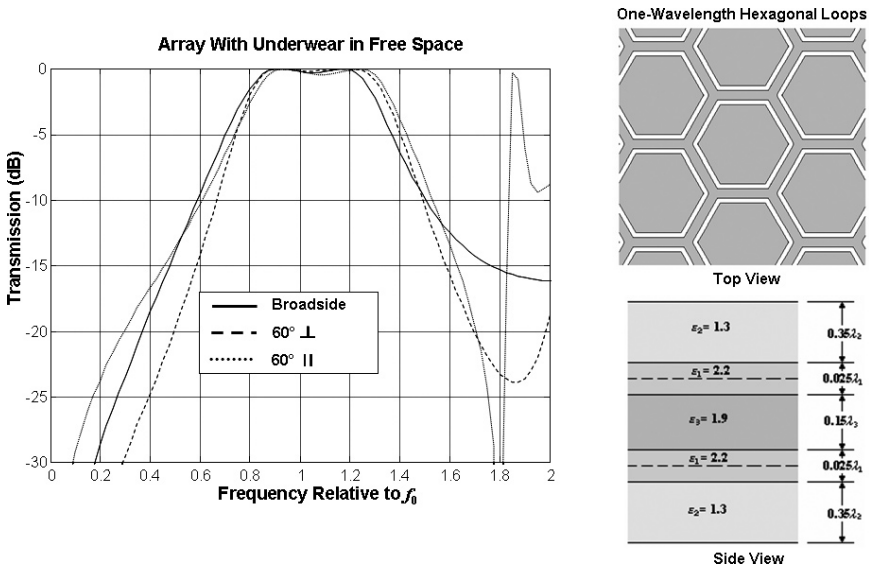


FIGURE 56-20 Transmission curves of cascaded FSS of hexagonal slot elements with dielectric slabs

Cascading FSS

While the bandwidth achieved by the FSS shown in Figure 56-19 is significant, there is (as with all single-layer FSS) essentially only one fundamental resonance. Multiple cascaded arrays can be used to shape the response curve so that it has more of a flattened top as well as a sharper roll-off. Analysis of multilayered FSS has been mentioned earlier in Section 56.2, and an in-depth discussion can be found in the reference section.²⁷ The choices in the number of layers, the separation between the arrays, and the dielectric properties of the material separating them all have effects on FSS performance. Figure 56-20 shows the transmission response for a FSS of two cascaded arrays. The arrays shown in Figure 56-20 are the same as the ones in Figures 56-18 and 56-19. Clearly, the addition of the second array has the desired effect of flattening the top of the transmission curve and sharpening the roll-off on both the lower and upper frequencies.

56.5 OTHER APPLICATIONS

Frequency selective surfaces are used in a variety of applications beyond those described in the preceding material. This final section will briefly touch on a few of those applications, in particular the ones that have been enabled by technological advances not originally developed for frequency selective surface applications. Undoubtedly, as time proceeds, more and more advanced applications will make use of frequency selective surfaces.

Reconfigurable arrays have received considerable interest at the dawn of the 21st century. A reconfigurable array has some means to change the performance of the array as needed for an application, which means that antennas or scattering arrays no longer need be considered stationary, but that they can be adapted to meet evolving needs. For example, there are some cases where an array is designed for wide bandwidth applications.⁵²

In others,⁵³ the array is reconfigured using magnetic field actuation, while in still others,⁵⁴ a static electric field is used to modify the array properties. Varactor diodes have also been used to tune the array performance.⁵⁵ One design method utilizes optimization techniques presented for a reconfigurable array without a specific implementation technology (for example, a type of switch).⁵⁶

Generally, optimization methods have been used more and more for designing frequency selective surfaces that would not have been developed using traditional intuition and knowledge of electromagnetics. In one case⁵⁶ a genetic algorithm is used, while in another a particle swarm optimization approach is used.⁵⁷ These optimization techniques typically require multiple incarnations of the array during the design procedure, and therefore efficient computational models are useful. For example,⁵⁸ a model-order reduction procedure is used to parameterize the response of the surface. Undoubtedly, advancing computational design methods will generate a variety of applications that provide niche solutions to challenging problems.

Although the majority of this chapter dealt with designs suitable for microwave and millimeter applications, significant work on shorter wavelength applications has been published.^{22,59} A significant application for periodic array surfaces is in forming reflectarrays. The literature is extensive^{60–64} on this topic, and the reader is encouraged to consult it, among other sources, for designs utilizing patch antennas and designing utilizing slots.⁶⁵

ACKNOWLEDGMENTS

The authors would like to thank Ben Munk and Leo Kempel for their valuable insight and unending support.

REFERENCES

1. B. Munk, R. Taylor, T. Durham, W. Crosswell, B. Pigon, R. Boozier, S. Brown, M. Jones, J. Pryor, S. Ortiz, J. Rawnick, K. Krebs, M. Vanstrum, G. Gothard, and D. Weibelt, "A Low-Profile Broadband Phased Array Antenna," *IEEE Antennas. Prop. Symp.* (June 22–27, 2003): 448–451.
2. J. F. McCann, R. J. Marhefka, and B. A. Munk, "Developments on the Design of a Large Bandwidth Array of Slot Elements with Wide Scan Angle," *IEEE Antennas. Prop. Symp.* (July 9–14, 2006).
3. P. Munk, "On Arrays That Maintain Superior CP and Constant Scan Impedance for Large Scan Angles," *IEEE Trans. Antennas Propagat.*, vol. 52 (February 2003): 322–330.
4. W. Rotman, "Plasma Simulation By Artificial Dielectrics and Parallel-Plate Media," *IEEE Trans. Antennas Propagat.*, vol. 10 (January 1962): 82–95.
5. I. J. Bahl and K. C. Gupta, "A Leaky-Wave Antenna Using An Artificial Dielectric Medium," *IEEE Trans. Antennas Propagat.*, vol. 21 (January 1974): 119–122.
6. I. J. Bahl and K.C. Gupta, "Radiation From A Dielectric-Artificial Dielectric Slab," *IEEE Trans. Antennas Propagat.*, vol. 23 (January 1976): 73–76.
7. C. Kyriazidou, R. Diaz, and N. Alexopoulos, "Novel Material with Narrow-Band Transparency Window in the Bulk," *IEEE Trans. Antennas Propagat.*, vol. 48 (January 2000): 107–116.
8. K. A. Zaki and A. R. Neureuther, "Scattering From A Perfectly Conducting Surface with A Sinusoidal Height Profile: TE polarization," *IEEE Trans. Antennas Propagat.*, vol. 19 (March 1971): 208–214.
9. Z. Nazarchuk and O. Ovsyannikov, "Diffraction of Electromagnetic Waves on Multielement Diffraction Grating," *IEEE Trans. Antennas Propagat.*, vol. 45 (January 1997): 15–19.
10. Y. Rahmat-Samii and A. N. Tulinseff, "Diffraction Analysis of Frequency Selective Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. 41 (April 1993): 476–487.

11. V. D. Agrawal and W. A. Imbraiale, "Design of a Dichroic Cassegrain Subreflector," *IEEE Trans. Antennas Propagat.*, vol. 27 (July 1979): 466–473.
12. T. Y. Otoshi and M. M. Franco, "Dual Passband Dichroic Plate for X-Band," *IEEE Trans. Antennas Propagat.*, vol. 40 (October 1992): 1238–1245.
13. L. W. Epp, P. H. Stanton, R. E. Jorgenson, and R. Mittra, "Experimental Verification of an Integral Equation Solution for a Thin-Walled Dielectric Plate with Cross-Shaped Holes," *IEEE Trans. Antennas Propagat.*, vol. 42 (June 1994): 878–882.
14. A. Monorchio, P. Grassi, and G. Manara, "A Hybrid Mode-Matching Finite Elements Approach to the Analysis of Thick Dichroic Screens with Arbitrary Shaped Apertures," *IEEE Antennas Wireless Propagat Ltrs.*, vol. 1 (January 2002): 120–123.
15. P. Besso, M. Bozzi, L. Perregrini, L.S. Drioli, and W. Nickerson, "Deep-Space Antenna for Rosetta Mission: Design and Testing of the S/X band Dichroic Mirror," *IEEE Trans. Antennas Propagat.*, vol. 51 (March 2003): 116, 388–394.
16. P. Franchi and R. Mailloux, "Theoretical and Experimental Study of Metal Grid Angular Filters for Sidelobe Suppression," *IEEE Trans. Antennas Propagat.*, vol. 31 (May 1983): 445–450.
17. Y. J. Lee, S. H. Jeong, W. S. Parck, J. S. Yun, and S. I. Jeon, "Multilayer Spatial Angular Filter with Airgap Tuners to Suppress Grating Lobes of 4x/1 Array Antenna," *Electronic Letters*, vol. 39 (January 2003): 15–17.
18. Y. Boucher, T. Fessant, and A. V. Uskov, "Angular and Spectral Filtering in Tilted-Grating Structures," *IEE Proc. Optoelec.*, vol. 146 (February 1999): 35–38.
19. R. Mailloux, "Synthesis of Spatial Filters with Chebyshev Characteristics," *IEEE Trans. Antennas Propagat.*, vol. 24 (March 1976): 174–181.
20. R. Giusto and P. De Vincenti, "Phase-Only Optimization for the Generation of Wide Deterministic Nulls in the Radiation Pattern of Phased Arrays," *IEEE Trans. Antennas Propagat.*, vol. 31 (September 1983): 814–817.
21. C. Zhou, F. Haber, and Q. Shi, "On Spatial Filtering for Angle-of-Arrival Estimation in a Scattering Environment by Eigendecomposition-Based Methods," *IEEE Trans. Antennas Propagat.*, vol. 38 (June 1990): 931–934.
22. E. Topsakal and J. Volakis, "Frequency Selective Volumes for Optical Spatial Filters," *IEEE Antennas Wireless Propagat Ltrs.*, vol. 3 (January 2004): 236–238.
23. T. Cwik, R. Mittra, K. C. Long, and T. K. Wu, "Frequency Selective Screens," *IEEE Antennas Propagat. Soc. Newsletter* (April 1987): 6–10.
24. R. Mittra, C. Chan, and T. Cwik, "Techniques for Analyzing Frequency Selective Surfaces—A Review," *Proc. IEEE*, vol. 76 (December 1988): 1593–1615.
25. T. K. Wu, (ed.), *Frequency Selective Surface and Grid Arrays* (New York: Wiley, 1995).
26. J. C. Vardaxoglou, *Frequency Selective Surfaces: Analysis and Design* (Taunton, England: Research Studies Press, 1997).
27. B. A. Munk, *Frequency Selective Surfaces* (New York: Wiley, 2000).
28. B. A. Munk, *Finite Antenna Arrays and FSS* (New York: Wiley, 2003).
29. T. Cwik and R. Mittra, "The Effects of the Truncation and Curvature of Periodic Surfaces: A Strip Grating," *IEEE Trans. Antennas Propagat.*, vol. 36 (May 1988): 612–622.
30. J. M. Usoff and B. A. Munk, "Edge Effects of Truncated Periodic Surfaces of Thin Wire Elements," *IEEE Trans. Antennas Propagat.*, vol. 42 (July 1994): 946–953.
31. B. A. Munk, D. S. Janning, J. B. Pryor, and R. J. Marhefka, "Scattering from Surface Waves on Finite FSS," *IEEE Trans. Antennas Propagat.*, vol. 49 (December 2001): 1782–1793.
32. B. A. Munk and G. A. Burrell, "Plane-Wave Expansion for Arrays of Arbitrarily Oriented Piecewise Linear Elements and Its Application in Determining the Impedance of a Single Linear Antenna in a Lossy Half-Space," *IEEE Trans. Antennas Propagat.*, vol. 27 (May 1979): 331–343.
33. L. Nord, "Sensitivity of Wave Propagation Characteristics to Manufacturing Defects in Frequency Selective Radomes," MS thesis, Lund University, 2005.

34. S. Schneider, "On the Scattering Properties of Super Dense Periodic Surfaces," Ph.D. Dissertation, The Ohio State University, 1992.
35. R. C. Hansen, *Microwave Scanning Antennas*, vol. 2 (New York: Academic Press, 1966): 203–208.
36. R. F. Harrington, *Time-Harmonic Electromagnetic Fields* (New York: McGraw-Hill, 1961): 164–171.
37. B. A. Munk and T. W. Kornbau, "On Stabilization of the Bandwidth of a Dichroic Surface by the Use of a Dielectric Slab," *Electromagnetics*, vol. 5 (1985): 349–373.
38. A. Mackay, "Proof of Polarization Independence and Non-Existence of Crosspolar Terms for Targets Presenting n -Fold ($n > 2$) Rotational Symmetry with Special Reference to Frequency Selective Surfaces," *Electronics Letters*, vol. 25, no. 24 (November 1989): 1624–1625.
39. C. Kittel, *Introduction to Solid State Physics* (New York: Wiley, 1986).
40. T. W. Kornbau, "Analysis of Periodic Arrays of Rotated Linear Dipoles, Rotated Crossed Dipoles, and of Biplanar Dipole Arrays in Dielectric," Ph.D. dissertation, The Ohio State University, 1984.
41. S. W. Schneider and B. A. Munk, "The Scattering Properties of 'Super Dense' Arrays of Dipoles," *IEEE Trans. Antennas Propagat.*, vol. 42, no. 2 (April 1994): 463–472.
42. S. B. Savia and E. A. Parker, "Equivalent Circuit Model for Superdense Linear Dipole FSS," *IEE Proc. Microwave Ant. and Prop.*, vol. 150 (February 2003).
43. R. Orta, P. Savi, and R. Tascone, "Numerical Green's Function Technique for the Analysis of Screen Perforated by Multiple Connected Apertures," *IEEE Trans. Antennas Propagat.*, vol. 44, no. 6 (June 1996): 765–776.
44. R. A. Hill and B. A. Munk, "The Effect of Perturbating a Frequency-Selective Surface and Its Relation to the Design of a Dual-Band Surface," *IEEE Trans. Antennas Propagat.*, vol. 44, no. 3 (March 1996): 368–374.
45. E. A. Parker, S. M. A. Hamdy, and R. J. Langley, "Arrays of Concentric Rings as Frequency-Selective Surfaces," *Electronics Letters*, vol. 17 (November 1981): 880–881.
46. F. Huang, J. C. Batchelor, and E. A. Parker, "Interweaved Convolved Element Frequency Selective Surfaces with Wide Bandwidths," *Electronics Letters*, vol. 42, no. 14 (July 6, 2006).
47. R. M. Foster, "A Reactance Theorem," *Bell System Technical Journal*, vol. 3 (April 1924): 259–267.
48. R. J. Langley and E. A. Parker, "Equivalent Circuit Model for Arrays of Square Loops," *Electronics Letters*, vol. 18 (1982): 294–296.
49. E. A. Parker and S. M. A. Hamdy, "Rings as Elements for Frequency Selective Surfaces," *Electronics Letters*, vol. 17 (1981): 612–614.
50. J. F. McCann, "On the Design of Large Bandwidth Arrays of Slot Elements with Wide Scan Angle Capabilities," Master's Thesis, The Ohio State University, 2006.
51. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd Ed. (New York: Wiley & Sons, 1998).
52. Y. E. Erdemli, K. Sertel, R. A. Gilbert, D. E. Wright, and J. L. Volakis, "Frequency-Selective Surfaces to Enhance Performance of Broad-Band Reconfigurable Arrays," *IEEE Trans. Antennas Propagat.*, vol. 50 (December 2002): 1716–1724.
53. J. Zendejas, J. P. Gianvittorio, Y. Rahmat-Samii, and J. Judy, "Magnetic MEMS Reconfigurable Frequency-Selective Surfaces," *J. MEMS*, vol. 15 (June 2006): 613–623.
54. B. Schoelinner, A. Abbaspour-Tamijani, L. C. Kempel, and G. M. Rebeiz, "Switchable Low-Loss RF MEMS Ka-Band Frequency-Selective Surface," *IEEE Trans. Microwave Theory Tech.*, vol. 52 (November 2004): 2474–2481.
55. C. Mias, "Varactor-Tunable Frequency Selective Surface with Resistive-Lumped-Element Biasing Grids," *IEEE Microwave Wireless Comp. Ltrs.*, vol. 15 (September 2005): 570–572.
56. J. Bossard, D. H. Werner, T. S. Mayer, and R. P. Drupp, "A Novel Design Methodology for Reconfigurable Frequency Selective Surfaces Using Genetic Algorithms," *IEEE Trans. Antennas Propagat.*, vol. 53 (April 2005): 1390–1400.

57. S. Genovesi, R. Mittra, A. Monorchio, and G. Manara, "Particle Swarm Optimization for the Design of Frequency Selective Surfaces," *IEEE Antennas Wireless Propagat Ltrs.*, vol. 5 (December 2006): 277–279.
58. D. S. Weile and E. Michielssen, "Analysis of Frequency Selective Surfaces Using Two-Parameter Generalized Rational Krylov Model-Order Reduction," *IEEE Trans. Antennas Propagat.*, vol. 49 (November 2001): 1539–1549.
59. J. A. Bossard, D. H. Werner, T. S. Mayer, J. A. Smith, Y. U. Tang, R. P. Drupp, and Li Ling, "The Design and Fabrication of Planar Multiband Metallodielectric Frequency Selective Surfaces for Infrared Applications," *IEEE Trans. Antennas Propagat.*, vol. 54 (April 2006): 1265–1276.
60. R.D. Javor, Wu Xiao-Dong, and K. Chang, "Design and Performance of a Microstrip Reflectarray Antenna," *IEEE Trans. Antennas Propagat.*, vol. 43 (September 1995): 932–939.
61. D.M. Pozar, S.D. Targonski, and R. Pokuls, "A Shaped-Beam Microstrip Patch Reflectarray," *IEEE Trans. Antennas Propagat.*, vol. 47 (July 1999): 1167–1173.
62. M. E. Bialkowski and H. J. Song, "Investigations into a Power-Combining Structure Using a Reflectarray of Dual-Feed Aperture-Coupled Microstrip Patch Antennas," *IEEE Trans. Antennas Propagat.*, vol. 50 (June 2002): 841–849.
63. R. Leberer and W. Menzel, "A Dual Planar Reflectarray with Synthesized Phase and Amplitude Distribution," *IEEE Trans. Antennas Propagat.*, vol. 53 (November 2005): 3534–3539.
64. M. R. Chaharmir, J. Shaker, M. Cuhaci, and A-R Sebak, "Novel Photonically-Controlled Reflectarray Antenna," *IEEE Trans. Antennas Propagat.*, vol. 54 (April 2006): 1134–1141.
65. M. R. Chaharmir, J. Shaker, M. Cuhaci, and A-R Sebak, "Reflectarray with Variable Slots on Ground Plane," *IEE Proc. Microwaves, Antennas, Propagat.*, vol. 150 (December 2003): 436–439.

Chapter 57

Propagation Models and Antennas for MIMO

Werner Wiesbeck
Christiane Kuhnert

University of Karlsruhe (TH)

CONTENTS

57.1	BASIC MIMO IDEAS	57-2
57.2	THE MIMO PROPAGATION CHANNEL	57-3
57.3	MIMO CHANNEL MODELING	57-5
57.4	ANTENNAS IN MIMO SYSTEMS	57-10
57.5	MIMO SYSTEM ASPECTS	57-15

57.1 BASIC MIMO IDEAS

Recently, the employment of antenna arrays at both sides of the link has become popular in both radar and communications. These so-called Multiple Input Multiple Output (MIMO) systems involve the combination of antenna arrays, sophisticated signal processing, and wave propagation. For radar systems, the main applications are interference reduction and beamforming, in order to increase the resolution and the quality of the signals. In communication systems the aims are similar, but here the enhancement of the spectral efficiency is an additional, very important topic. The idea was first published by Winters¹ in 1987, but not much attention was paid at that time. In 1998, Foschini et al² showed that MIMO systems reach much higher spectral efficiencies than today's communication systems. Much research has since been devoted to the field of MIMO.

Wireless systems operate via the transmission of electromagnetic waves. Various propagation effects, such as the reflection of electromagnetic waves at walls, occur depending on the environment. Various means by which the waves reach the receiver are referred to as propagation paths. Usually, waves are received through various paths. This phenomenon is called *multipath propagation*. Transmission systems suffer from multipath, because the waves coming from different directions at different times interfere. This causes signal distortions called *fading*.

Multipath propagation is not a drawback, but is beneficial for MIMO, which makes use of the spatial dimension of the propagation channel. With each antenna, the energy from all waves is collected and the incoming streams are processed for coherent summation. In the case of uncorrelated reception, fading can be overcome. If two or more antennas at the receiver are spaced far enough apart that their fading envelopes are uncorrelated, the probability of system failure is dramatically reduced. In case of weak reception at one antenna, the other antenna can still provide a sufficient signal. The same effect can be utilized at the transmitter. The total transmit power is distributed between the transmit antennas and transmitted from more than one antenna. One major question is figuring out the ways that multiple antennas should be positioned for uncorrelated reception.

With multiple antennas, the spatial structure of the channel may be well exploited by using combinations of diversity and digital beamforming techniques. It is possible to transmit several datastreams simultaneously, in parallel, and at the same frequency as shown in Figure 57-1.

Figure 57-1 demonstrates how MIMO exploits the multipath structure. It shows the top view of an urban scenario with several buildings. The dominant paths between a transmitter and a receiver, which are indicated as rays, experience various phenomena such as reflection or diffraction. By employing antenna arrays at both transmit and receive, the antenna patterns can be adapted to the dominant propagation paths. If this beamforming is done in the digital domain, two (or more) different antenna patterns can be created, such that each antenna pattern focuses on different propagation paths. One datastream is transmitted by each antenna pattern. Hence multiple datastreams can be transmitted in parallel at the same frequency—spatially separated via different propagation paths—leading to a tremendous spectral efficiency.

MIMO systems exploit the multipath structure of the propagation channel by adapting the antennas to the current situation. The exploitation of the spatial structure of the propagation channel leads to an efficient use of the basic resources time, spectrum, code, and space.

A discussion of the MIMO propagation channel in Section 57.2 is followed by a description of channel modeling approaches (Section 57.3). Section 57.4 deals with antennas for MIMO systems and, finally, future prospects are given in Section 57.5.

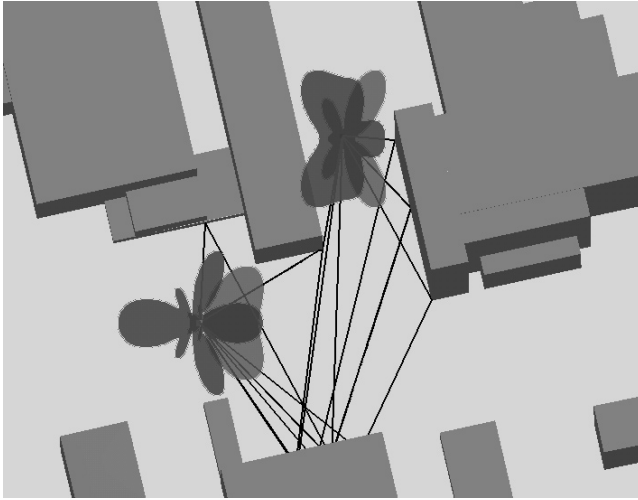


FIGURE 57-1 MIMO exploiting the multipath structure in an urban scenario

57.2 THE MIMO PROPAGATION CHANNEL

The propagation channel is the heart of any wireless system. It is the principal contributor to many of the problems and limitations that beset mobile radio systems.

Spatial Properties and Angular Spread

A signal propagating through the wireless channel is affected by phenomena such as scattering, reflection, refraction, and diffraction of the radiated energy. Due to the large number of different propagation mechanisms, a large number of different paths usually exist between the transmitter and receiver that the signal has propagated on. These phenomena give rise to fading and path-loss.

There are two different kinds of fading. *Slow fading* arises due to shadowing effects of, for example, buildings, trees, and terrain variations, which cause the signal to fluctuate around the mean path-loss. *Fast fading* is caused by the addition of multiple paths, arriving at slightly different times, with different amplitudes and with different phases. These signal paths sometimes add constructively and sometimes destructively at the receiver, and may cause a variation in the power level of up to 20-30 dB over a distance of half a wavelength. The mean path-loss depends on the transmission distance and the environment. The loss in free space is 20 dB/decade, but in a more complex environment a loss of 40 dB/decade is common.

Due to the multipath, a signal will arrive at the receiver by different paths. Since these paths have different lengths and angles of arrival, there will be a spread of the signal in time and direction. The spread in direction is referred to as angular spread, a fundamental means to describe the spatial properties of transmission channels. The power azimuth spectrum describes the spatial wave propagation properties found at a transmit or receive antenna.

It is a statistical distribution function, modeling the power distribution versus the azimuth angle. A similar spectrum can be defined for the elevation angle.

For MIMO systems, the directions of departure and arrival of the paths are crucial, because different signals are separated by these angles. The more directions that exist, the better it is for MIMO. In indoor scenarios there are usually many reflecting objects between the transmitter and receiver, resulting in a large angular spread. Depending on the environment, the angular spread may vary in outdoor scenarios. Rural environments lack interacting objects, which means that the angular spread is much smaller than in urban environments with many buildings. Satellite links are usually very directive, leading to a very small angular spread.

Channel Transfer Matrix

The small scale variations of a radio signal can be directly related to the impulse response of the radio channel. The impulse response is a wideband channel characterization and contains all the information necessary to analyze any type of radio transmission. This stems from the fact that a radio channel may be modeled as a linear filter with a time-varying impulse response $h(t, \tau)$, where the temporal variation (described by the variable t) is due to spatial receiver motion or time-varying propagation conditions (for example, moving obstacles). Alternatively, the radio channel can be characterized by the Fourier transform of the impulse response, for example, the time-varying channel transfer function $h(f, t)$. It is defined as the ratio of the receive voltage V_R to the transmit voltage V_T (voltages at the antenna feeds).

$$h(f, t) = \frac{V_R(f, t)}{V_T(f, t)}$$

Under certain conditions, the equivalent baseband impulse response can be written as

$$h(\tau, t) = \sum_{i=1}^{P(t)} b_i(t) e^{j\phi_i(t)} \delta(\tau - \tau_i)$$

This is very intuitive. The received signal for an impulse $\delta(t)$ exciting the channel consists of a series of attenuated (b_i), phase-shifted (ϕ_i), and time-delayed ($\tau - \tau_i$) replicas of the signal, each corresponding to one of the P paths. However, in reality the channel is always band-limited. Thus the Dirac impulse has to be replaced by some filter function characterizing the finite bandwidth of the channel. If the multipath components arrive at the receiver almost at the same time compared to the symbol duration, the channel is called *flat fading channel*. In this case, the channel transfer function merges into a complex channel coefficient h describing the attenuation and phase shift of the channel.

Figure 57-2 shows the radio frequency link of a MIMO system. It consists of several transmit antennas, the physical channel, and several receive antennas. Each transmitted signal is received by all receive antennas. The channel is described by the complex channel coefficient for the flat fading case. For each pair of transmit (T, j) and receive (R, i) antennas, the channel coefficient is

$$h_{ij}(t) = \frac{V_{R,i}(t)}{V_{T,j}(t)}$$

the ratio of the complex voltage at the receive antenna to the complex voltage at the transmit antenna.

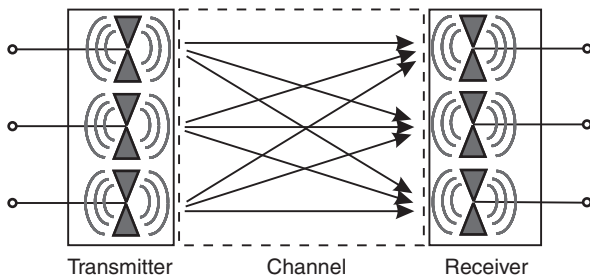


FIGURE 57-2 MIMO system

For a MIMO system with M transmit and N receive antennas, the channel is described by the channel matrix

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & & \vdots \\ & & \ddots & \vdots \\ h_{N1} & \cdots & \cdots & h_{NM} \end{pmatrix}$$

which contains the channel coefficients linking all different pairs of transmit and receive antennas. The channel matrix \mathbf{H} is the fundamental means of describing the MIMO transmission channel. It serves as the input for the calculation of the spectral efficiency, which is often referred to as the *MIMO capacity*.

In a MIMO system there is not a single transmit signal, but a whole transmit signal vector \mathbf{s} whose elements describe the signals transmitted by the single antennas. Hence the receive signal is also a vector \mathbf{y} . The whole MIMO system is described by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

where \mathbf{n} is an N -dimensional additive Gaussian noise vector.

57.3 MIMO CHANNEL MODELING

For system design, testing, antenna development, and evaluation, a model of the propagation channel is necessary. Early classical models, which were developed for narrowband transmission systems, only provided information about the signal power level distribution and Doppler shifts of the received signals. An example is the Hata-model,³ which is widely used in GSM network planning. It predicts the log-normal path-loss and thereby provides an estimate of the mean signal level.

As wireless systems became more complex and more accurate models were required, additional concepts such as time delay and angle of arrival were incorporated into channel models. For MIMO systems the spatial structure, temporal structure, the joint spatial and temporal structure, and polarization behavior have to be modeled as well. There are many studies addressing the properties of the MIMO propagation channel. These efforts have resulted in realistic characterizations of the propagation channel for different environments, for example, indoor or outdoor.

Stochastic Models and the Extrapolation Method

Stochastic models reflect time-variant channel behavior and have the advantage of short simulation times. Most channel models consider only channel impulse responses without any more detailed multipath information. MIMO systems require the consideration of the radiation directions of the multipath components at the transmitter and their angles of arrival at the receiver. Zwick et al⁴ have introduced a stochastic channel model based on physical wave propagation, which allows a fully polarimetric wideband description of the multipath channel including path directions. Each multipath component is characterized by its transfer matrix (including loss), delay, direction of arrival, and departure. The appearance and disappearance of multipath components over time is modeled as a birth and death process (a marked Poisson process), which enables the correct modeling of spatial and temporal correlations. In each modeling step, path properties change according to the motion of the transmitter and receiver. The changing delay times of propagation paths yield a realistic Doppler behavior for the channel. Data sets, which are required for the statistical evaluation of the parameters of the model have been obtained by ray tracing simulations and narrow-band, wideband, and directional channel measurements. The azimuthal power profile of the stochastic channel model is composed of several Laplacian functions, each modeling a cluster of scatterers. The elevation profile is also modeled by Laplacian functions for paths with a short delay time, changing to a sinusoidal function for paths with a long delay time. The channel model delivers the channel impulse response, the angle of departure of all paths at the transmitter, and the angle of arrival of all paths at the receiver for the SISO (system with one transmit and one receive antenna) case.

With knowledge of the channel for the SISO case, the MIMO channel matrix can be derived,⁵ as long as, first, the antenna spacings do not extend several wavelengths (for example, when the same plane waves impinge at all antenna elements), and second, the distances between transmitter and receiver and any obstacles in the channel are large enough to assume plane waves impinging at the arrays. With these assumptions the MIMO channel matrix results from the coherent addition of the impinging waves at the different antenna positions. The channel impulse responses, or in the flat-fading case the channel coefficients h , are derived by placing virtual antennas in the vicinity of the SISO transmit and receive antennas. The phase difference of the incident plane waves at the different antenna positions is described by $\Delta\varphi$ (illustrated in Figure 57-3), which is a function of the angle of arrival of the waves and the antenna positions:

$$\Delta\varphi = \frac{2\pi}{\lambda} ((x_{\text{SISO}} - x_{\text{vir}}) \cos\phi \sin\theta + (y_{\text{SISO}} - y_{\text{vir}}) \sin\phi \sin\theta + (z_{\text{SISO}} - z_{\text{vir}}) \cos\theta)$$

The difference of the amplitude of the plane wave at the positions of the virtual antennas is neglected; that is, the amplitude is assumed to be constant for all antenna positions. The result is the channel coefficient for all virtual receive and transmit antennas

$$h = \sum_{i=1}^{P(t)} b_{1,i} e^{j\varphi_{1,i}} e^{-j\Delta\varphi_{1,i}^{\text{Tx}}} e^{-j\Delta\varphi_{1,i}^{\text{Rx}}}$$

where $b_{1,i}$ is the attenuation of the i -th path in the SISO channel and $\varphi_{1,i}$ is its phase.

Deterministic Characterization

Ray tracing provides an adequate means to model the multipath wave propagation for MIMO channels. A channel model for outdoor environments has been introduced by

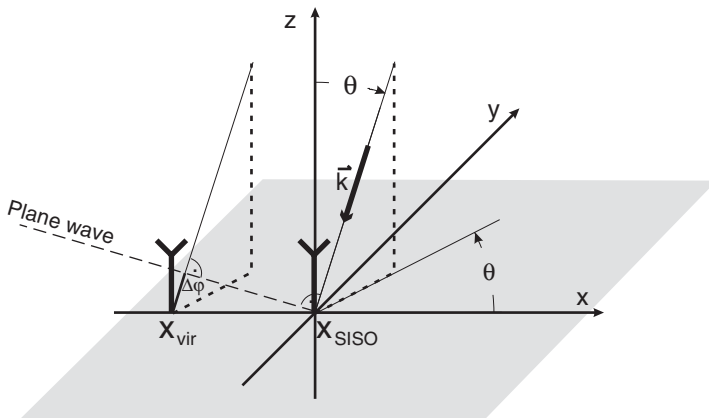


FIGURE 57-3 A plane wave is impinging on the array. The phase difference of the plane wave at the different antenna positions is $\Delta\varphi$.

Maurer et al,⁶⁻⁷ which allows for an accurate narrowband and wideband characterization of outdoor MIMO channels. The propagation phenomena taken into account are combinations of multiple reflections, diffractions and scattering from trees. Modified Fresnel reflection coefficients, which account for slightly rough surfaces, are used to model the reflections. Diffractions are described by the uniform geometrical theory of diffraction (UTD) and the corresponding coefficients for wedge diffraction. Scattering from trees is considered to be totally incoherent, i.e., no distinct specular component is present. This assumption holds if the wavelength is smaller than the dimensions of the randomly distributed leaves and branches.

Depending on the propagation phenomena, different approaches to ray tracing exist. In order to trace pure reflection paths, the method of image transmitters (image theory) is implemented. Since the proposed propagation model supports full 3D diffraction, Fermat's principle is used to determine the diffracted ray paths. For mixed paths, image theory and Fermat's principle are combined. As only single scattering from trees is taken into account, the scattering paths are defined by the position of transmitter and receiver and the position of the central point of the tiles into which the tree model is subdivided.

Figure 57-4 shows the result of the ray tracing for one snapshot of an urban road traffic scenario. The receiver and transmitter are positioned on two cars with a truck between them resulting in a non-line-of-sight condition. In order to keep track of the result, only the 30 strongest propagation paths are indicated (bright lines). It can be seen that part of the transmitted energy is diffracted around the truck. Since there are many buildings next to the road, the expected street canyon effect arises. Scattering from single trees contributes to the received signal as well.

For MIMO systems, ray tracing can be applied for the calculation of the channel linking each pair of transmit and receive antennas.⁸

Geometry-Based Stochastic Channel Modeling

Spatial properties of the radio channel—such as the directions of departure and arrival of paths—are crucial for MIMO systems. If MIMO is employed by several users, the

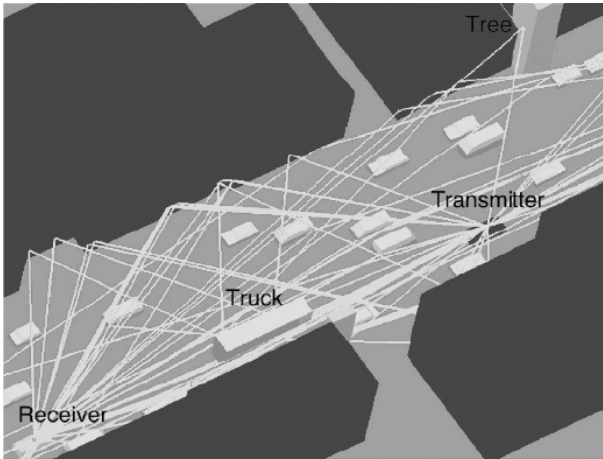


FIGURE 57-4 Ray tracing result

spatial properties of the channel experienced by adjacent users are similar. This means that for decreasing distance between neighboring links the correlation of the channel parameters should increase. Also these spatial dependencies should be incorporated in a channel model. Ray tracing is the best means to include all spatial properties. For MIMO systems, each channel coefficient of the channel transfer matrix has to be calculated for each user; thus, the calculation effort is extensive. The so-called geometry-based stochastic channel models provide a solution—the fast calculation time of stochastic models is combined with spatial properties and dependencies provided by ray tracing.

A geometry-based stochastic channel model has been introduced by Fügen et al.⁹ The propagation environment is randomly generated in pre-processing and stays fixed during the whole simulation. It is an urban environment consisting of buildings and streets, which are arranged as a Manhattan-like street grid. All buildings are placed along the streets. A number of fixed scatterers are distributed uniformly in the whole 3D simulation environment. The simulation environment may contain several MIMO transmitters and MIMO receivers. Each mobile station is located in a circular/elliptical cylinder that defines the local cluster (as shown in Figure 57-5). The cylinder serves as a search function—only scatterers within the cylinder are active scatterers and contribute to the channel transfer function. Scatterers are modeled as ideal conductive and rectangular plates. The single-bounced rays between the mobile station and the base station can be calculated according to the scattering models also applied in ray tracing.

To account for a time variant channel behavior, the channel model is combined with a mobility model. Mobile stations move linearly along the street and may change their direction at each intersection. When the mobile station MS moves to the point MS', a number of new scatterers contribute to the received signal, while at the same time some old scatterers fade out. The shape and the size of the local cluster is adjusted depending on the position of the mobile station. If the mobile station is located in a street canyon, the local cluster is modeled as an elliptical cylinder aligned parallel to the street. If the mobile station is positioned at an intersection, the local cluster is approximated by a circular cylinder.

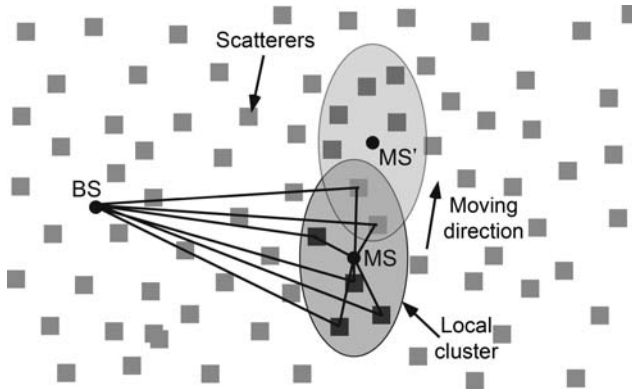


FIGURE 57-5 Scatterers and the local cluster

In macro-cellular environments, the street grid with adjacent buildings on both sides represents a waveguide. Interaction points in the surrounding of an intersection are placed as additional scatterers in a circular cylinder at each intersection, as shown in Figure 57-6. The rays propagate from the base station towards these scatterers. If the scatterer is visible from the mobile station, the scattered ray propagates as a guided wave along the street canyon towards the mobile station. This can be calculated by the 3D image theory.

To give a complete description of the wireless channel, models for far-distance clusters and line-of-sight are included as well.

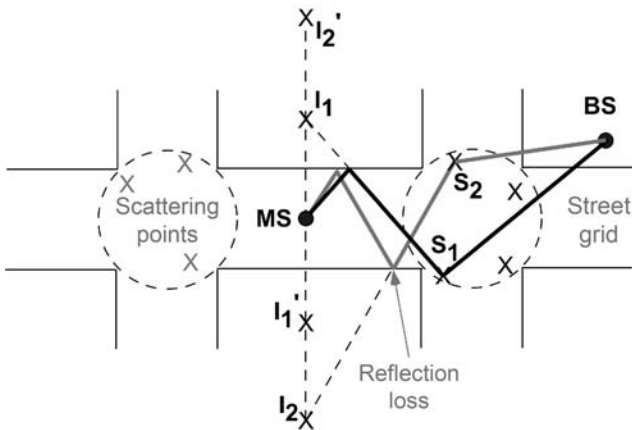


FIGURE 57-6 Scattering in a street canyon calculated by 3D image theory

57.4 ANTENNAS IN MIMO SYSTEMS

MIMO is associated with exploiting the spatial domain of the propagation channel and involves sophisticated signal processing. The employment of antenna arrays at both sides of the link is essential for MIMO, and finding feasible antenna configurations is an integral part of enabling the MIMO technology. The channel coefficients are influenced by the antenna properties. As the antennas are colocated in a MIMO array, mutual coupling effects may occur. All these effects should be considered when designing an antenna array for MIMO systems.

Antenna Evaluation

MIMO systems take advantage of the spatial properties of the propagation channel. Therefore, in the evaluation of MIMO antenna arrays, it is necessary to use a proper characterization of the spatial channel properties. Parameters, such as the angular spread of the waves impinging at the receiver, have a significant influence on the MIMO system performance. A meaningful conclusion from a comparison of MIMO antenna arrays can only be drawn if the spatial properties of a realistic propagation channel have been taken into account.

The combination of sophisticated signal processing and multiple antennas leads to an adaptation of the MIMO system to the multipath structure of the channel. In environments where MIMO systems can be applied, the propagation situation changes rapidly. Hence it is not sufficient to investigate the performance of MIMO antennas at one specific moment and use only one realization of the propagation channel. The performance of a MIMO array is determined by the ability of the array to adapt to the changing channel conditions. As such, antenna evaluation is not possible in a simple manner. Both the antenna arrays and the propagation channel should be treated together, with a statistical description, to take into account different channel realizations of a given propagation environment. To obtain statistically relevant performance data when comparing multi-antenna configurations, several hundred meters of measurement routes in several types of propagation environments would be needed for each prototype antenna. This is almost impossible in practice due to the large amount of measurements needed, but also due to restrictions imposed by the authorities on the usage of the frequency bands in which commercial communications networks are already operating. Channel models allow for simulations required to evaluate MIMO antennas.

For the development of MIMO antenna arrays, evaluation mechanisms are needed in order to draw comparisons between different antenna arrays. There are several antenna properties that can be assessed by classical antenna measures, for example, radiation pattern, antenna gain, self and mutual coupling impedances, half-power beamwidth, bandwidth, frequency range, and resonant frequency. If an antenna array is employed in a MIMO system, some more specific figures of merit that can take into account the propagation channel are required. These figures of merit are statistical measures, which describe the antenna array performance for specific propagation channels. The following is an overview of MIMO-specific evaluation criteria for antenna arrays.

As MIMO systems are known for their capacity enhancement for future mobile communications, the instantaneous channel capacity is the most important measure for the evaluation of MIMO systems. It is described in the following paragraph. There are two main properties that determine the capacity of a MIMO system: correlation properties and the power level or efficiency in terms of power. Hence evaluation criteria are needed both for the correlation properties and for the efficiency in terms of power which will be introduced in subsequent paragraphs.

The channel capacity depends on the number of transmit and receive antennas M and N , the channel matrix \mathbf{H} , and the signal-to-noise ratio (SNR). The channel matrix is often

normalized so that it is independent of the channel attenuation. \mathbf{H}_F is the channel matrix that is normalized with the Frobenius matrix norm, so that the mean attenuation of each channel matrix trace $(\mathbf{H}_F \mathbf{H}_F^*)/MN$ is equal to one. The capacity is expressed as a function of the SNR at the receiver. For the sake of simplicity only MIMO systems without channel state information at the transmitter are considered in the following. With this assumption, the transmit power is equally (but not optimally) spread among the transmit antennas. The instantaneous channel capacity of a MIMO system in the presence of spatially uncorrelated Gaussian distributed noise can be calculated by

$$C = \log_2 \left(\det \left(\mathbf{I} + \frac{\text{SNR}}{M} \mathbf{H}_F \mathbf{H}_F^* \right) \right)$$

where I is the identity matrix and $(\cdot)^*$ denotes the complex conjugate transposition. The channel attenuation, which is included in the channel matrix, has to be expressed in the SNR when normalizing the channel matrix. The attenuation is influenced by the antennas and the radio channel. With this normalization, the influence of the correlation properties on the capacity becomes visible, but any interrelation between the SNR and the correlation properties of \mathbf{H} is neglected. As the antennas have an impact on both the SNR and the correlation properties, this normalization does not predict the behavior of different antenna arrays properly. In real systems, the effect of antennas is included in the channel matrix and the normalization of the channel matrix eliminates some of the antenna effects. Therefore another approach is needed to preserve the antenna effects, allowing for a fair comparison of MIMO antenna arrays. If \mathbf{H} is not normalized, as in if the path loss and the gains of the single antenna elements are included in \mathbf{H} , the capacity of a MIMO system can be written as

$$C = \log_2 \left(\det \left(\mathbf{I} + \frac{P_T}{\sigma_n^2} \mathbf{H} \mathbf{H}^* \right) \right)$$

This equation expresses the capacity as a function of the transmit power P_T . The attenuation of the transmission link, influenced by the antennas and the radio channel, is taken into account. σ^2 is the noise power. This formula allows for a comparison of different MIMO systems, including the influence of the transmission gain and the SNR. Without normalization the capacity is an appropriate figure of merit for the comparison of MIMO systems employing different antenna arrays.

The correlation properties of \mathbf{H} influence the capacity. The different definitions commonly used are the complex signal (ρ_s), envelope (ρ_e), and the power correlation (ρ_p) coefficients. One can express the complex signal correlation coefficient as

$$\rho_s = \frac{E\{h_1 h_2^*\}}{\sqrt{E\{h_1 h_1^*\} E\{h_2 h_2^*\}}}$$

where h_1 and h_2 are complex channel coefficients and E is the expectation operator. h_1 and h_2 represent any two channel coefficients of the channel matrix \mathbf{H} . The definition of the envelope and the power correlation coefficients are

$$\rho_e = \frac{E\{|h_1| \cdot |h_2|\}}{\sqrt{E\{|h_1| \cdot |h_1|\} E\{|h_2| \cdot |h_2|\}}}$$

$$\rho_p = \frac{E\{|h_1|^2 \cdot |h_2|^2\}}{\sqrt{E\{|h_1|^2 \cdot |h_1|^2\} E\{|h_2|^2 \cdot |h_2|^2\}}}$$

In literature, the most common reference to MIMO performance is in terms of the power and envelope correlation.¹⁰ It is a widely held assumption that $|\rho_s|^2 \approx \rho_e$. It has been both experimentally and analytically shown that $\rho_e = \rho_p$ for all practical purposes. The power correlation coefficients can be calculated for MIMO systems with different antenna arrays, leading to a comparison of antenna arrays for MIMO. The number of correlation coefficients between all elements h_{ij} in \mathbf{H} is N^2M^2 , and thus it is a large effort to assess the correlation properties. It is difficult to show the direct relationship between the capacity distribution and the correlation properties. Correlations also play an important role in diversity systems with multiple antennas at one side of the link only.

The power level is the second important property influencing the capacity. To assess single antennas in an array, the mean effective gain (MEG) can be used.¹¹ The MEG is defined as the ratio of the mean received power of an antenna under test to the mean received power of a reference antenna, when both antennas are used in the same channel with the same transmit antenna. To ascertain that the MIMO system works properly, all antennas need to have an MEG that is approximately equal and as high as possible. If the antenna elements in a MIMO array do not have similar mean effective gains, a branch power imbalance may occur, deteriorating the system performance. The definition of the MEG can be extended to assess arrays. The mean effective array gain (MEAG) is the ratio of the mean received power of an array to the mean received power of a reference antenna in the same channel with the same transmit antenna. Mutual coupling among closely spaced antennas not only influences the signal flow and the correlation properties, but can also strongly reduce the efficiency in terms of power of an array. This is undesirable since most handheld devices are battery powered.

The power transmission gain is a figure of merit for the whole MIMO link, including the antenna arrays at the transmitter, the propagation channel, and the antenna arrays at the receiver. The power transmission gain is the ratio of the power received at the signal drain to the power fed into the transmit antennas. The latter is not equal to the power which is radiated from the transmit antennas if the efficiency of the transmit array is not 100%. Conclusions on the performance of the arrays in terms of power can be drawn by comparing the power transmission gain of MIMO systems with different arrays in the same channel.¹²

The channel capacity measured in a reverberation chamber has been used as a figure of merit instead of quality measure for the antenna configurations.¹³ The measurements are repeatable, allowing for the comparison of different antenna arrays, although the environment does not necessarily correspond to real propagation channels.

Mutual Coupling and Radiation Efficiency

As mentioned before, the antennas have an impact on efficiency in terms of power. The mean effective gain, mean effective array gain, and power transmission gain are figures of merit for the efficiency of MIMO antenna arrays. The antenna configuration has a significant influence on the total received power. Mutual coupling effects found among the antenna elements in a MIMO array have an influence on power and capacity still under investigation. Seemingly conflicting reports on mutual coupling between antenna elements increasing¹⁴ or reducing¹² capacity might not be conflicting at all. Both effects are, in principle, plausible. On one hand, mutual coupling changes the individual antenna patterns. This creates diversity as each antenna “sees” different portions of the surrounding scatterers (pattern diversity). But the effect seems to be small unless the antennas are located very close to each other. On the other hand, mutual coupling may, by re-radiation of received power, result in higher spatial correlation between antenna signals; a possible cause for reduced capacity. However, correlation is not solely caused by mutual

coupling and depends on a number of other factors as well. Antennas in close proximity to each other and to lossy material, like human tissue, also suffer from reduced radiation efficiency. This is an effect which is intricate to measure, but it becomes extremely important when comparing the performance of an antenna array with that of a single antenna. One way to tackle this problem is the description of the MIMO communication link including transmit antennas, propagation channel, and receive antennas by scattering parameters.¹² Current findings indicate that spatial correlation has little effect in comparison with radiation efficiency.

Antenna Configurations

Due to the fact that the antennas and the propagation channel interact in MIMO systems, the antenna configuration has to be chosen carefully. The antenna type as well as the arrangement of the elements in a MIMO array strongly influences performance. Goals are a high efficiency in terms of power and a low correlation by exploiting various propagation paths. A major challenge is the integration of multiple antennas in small handheld devices as space for placing the antennas is extremely limited.

The following example uses half-wavelength dipole antennas to demonstrate how the antenna configuration affects MIMO capacity. In Figure 57-7, the capacity calculated without normalizing the channel matrix is shown for several antenna configurations. The propagation channel is modeled by a path-based indoor channel model. Various dipole arrangements are compared, for example, parallel dipoles, orthogonal dipoles, and other arrangements employing various orientations of the dipole elements. The capacity is shown with respect to the total size of an array, which is indicated in grey. This array size is very important for the integration of multiple antennas in handhelds. It is remarkable that for small array sizes an array with two elements outperforms arrays with three or four elements. Systems employing polarization diversity are robust against polarization mismatching, which is demonstrated by a rotation of the transmit array relative to the receive array. Three dipole antennas arranged as a triangle lead to the best result for small array sizes. This arrangement combines spatial, pattern, and polarization diversity, and thus the combination of different diversity techniques leads to capable solutions.

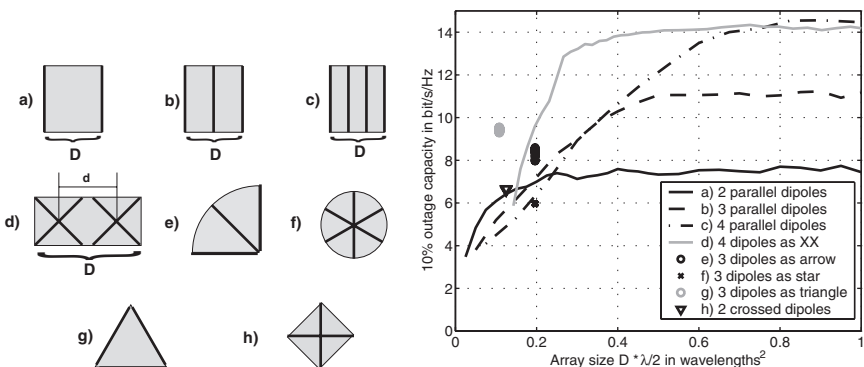


FIGURE 57-7 Arrangement of dipole antennas and MIMO capacity (after C. Waldschmidt et al¹² © IEEE 2004)

Integration of MIMO Antennas into Handhelds

Very compact antenna arrays are needed for the integration of antennas for MIMO into handheld devices. In this section, the integration of MIMO antenna arrays into handheld devices is demonstrated by two examples. The first example is a model of a portable device such as a mobile phone.¹⁵ It is equipped with three inverted-F antennas operating at a frequency of 2 GHz. The aim of the antenna configuration was to combine all the different diversity techniques described previously, such as pattern, spatial, and polarization diversity. The simulation model of the small handheld consists of a metallic block, representing the battery and the display of the device, and a PVC housing with a wall thickness of 2 mm, (as shown in Figure 57-8). The size of the housing is $55 \times 115 \times 27 \text{ mm}^3$ with the metallic block measuring $40 \times 80 \times 10 \text{ mm}^3$. Three antennas were mounted onto the metallic block, spatially separated and with different orientations. Usually inverted-F antennas require an infinite ground plane, which is not available in the small handheld. Thus the metallic block is effectively a finite ground plane, which influences the shape of the patterns and the mutual coupling impedances. The whole device was simulated with a standard EM code based on method of moments. The pattern of the coupled antenna system, as well as the mutual coupling and self-impedances of the antennas, were calculated and served as an input for the model of a MIMO transmission link.¹² The channel model is a stochastic, fully polarimetric, three-dimensional and path-based indoor channel model. On the other side of the link an antenna array setup, consisting of three half wavelength dipole antennas arranged as a triangle, was used.

For comparison, Table 57-1 presents figures of merit for the MIMO system mentioned previously, a much larger reference MIMO system with three vertical half-wavelength dipoles with half-wavelength spacings on each side of the link, and a single antenna system as reference. The correlation is very low for both MIMO cases. However, the capacity differs due to the fact that the mean effective gain is worse for the three small inverted

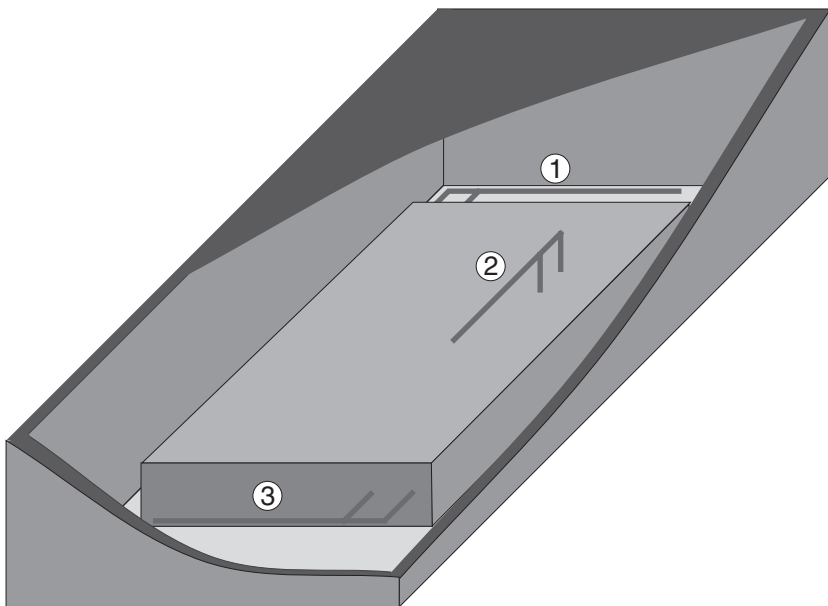


FIGURE 57-8 Model of a mobile phone with three inverted-F antennas (after C. Waldschmidt et al¹⁵ © IEE 2005)

TABLE 57-1 Performance Comparison of MIMO Systems¹⁵

	Power Correlation Coefficients	Mean Effective Gain (MEG)	10 Percent Outage Capacity
3 × 3, handheld and dipole triangle	all < 0.1	all antennas –2.5 dB	4.7 bit/s/Hz
3 × 3, parallel dipoles, half-wavelength spacing	all < 0.1	all antennas –1.2 dB	5.7 bit/s/Hz
1 × 1, dipole	–	0 dB	1.1 bit/s/Hz

F-antennas because of mutual coupling. Compared with a single antenna system, a huge capacity gain is possible, even if three antennas are integrated into a small handheld.

Another example¹⁶ discusses the case in which three antennas were integrated in a personal digital assistant measuring $80 \times 120 \times 15 \text{ mm}^3$. Each dual band antenna element covers the IEEE 802.11a/b frequency bands at 2.45 GHz and 5.2 GHz. Again the antennas are placed such that spatial, polarization, and pattern diversity are combined. Hence the power correlation is on the order of 0.1, which is sufficient for MIMO. This MIMO system outperforms a single antenna reference system in an indoor channel, even if one or two antennas are covered by the hand of a user.

When evaluating MIMO antenna configurations, both antennas and the propagation channel have to be treated simultaneously. Comparisons can be drawn employing special figures of merit for MIMO antenna arrays. The efficiency in terms of power has to be taken into account to assess arrays for MIMO. Several effects have an impact on the power transmission gain and on the MIMO performance, for example, the polarization of the antennas, the user of the handheld device, and mutual coupling effects between the antenna elements. These examples show that it is possible to effectively integrate several antennas into handhelds for MIMO operation. MIMO systems are realizable even in compact configurations.

57.5 MIMO SYSTEM ASPECTS

The potential of MIMO systems brings along several topics, not only concerning antennas and channel modeling. A question currently arising is the impact of MIMO on the complete RF transmission, including antennas, propagation channel, and RF front-ends.

RF and Digital Subsystems

Despite all the benefits of MIMO transmission, there is one major drawback: the increased complexity in both digital and analog systems. This complexity impacts both the cost of hardware and power consumption. Especially the analog part of the MIMO transceiver has to be designed carefully to avoid performance loss. Implementing MIMO techniques is a challenging task in conventional RF circuit design. The true test of the practicality of any communication system design is the performance of the system when taking into account the non-idealities of analog RF components. Although mutual interference between adjacent RF chains of a multi-antenna system causes additional performance degradation, a certain level of front-end integration will be required. MIMO techniques bring along some computational effort in the digital signal processing. The introduction of MIMO in future wireless systems leads to consequences for both RF and digital subsystems. A trade-off has to be found between quality and effort of implementation.

“Dirty” RF

The trade-off especially concerns the tough requirements for the quality of RF front-ends, as devices for the consumer market have to be affordable. Low power consumption, small size and low cost are important goals for the design of radio transceivers. However, these design goals partly contradict the demand for high performance. New problems arise in the practical implementation. Due to the finite accuracy of the analog hardware, the transmitted or received signals are threatened by several RF non-idealities such as jitter, non-linearities, phase noise, DC offset and I/Q imbalance. According to the conventional RF design paradigm, these effects have to be avoided by a proper system design leading to an increased power consumption, size or cost. Allowing the RF components to introduce analog impairments —“dirty effects”—and compensating the resulting effects in the digital domain will therefore be essential for achieving low production costs in the future.

When further integrating a complete MIMO transceiver on a single chip, these dirty effects reach into a new problem level, which requires a paradigm shift in the design of transceivers.¹⁷ “Dirty RF” means to design digital signal processing such that the system can cope with a new level of non-idealities, reducing the requirements of future RF subsystems. The “Dirty RF” design paradigm proposes to accept the RF non-idealities to a certain degree, and to compensate them digitally. Thus, power consumption, size, and cost of the transceiver can be kept at a low level.

REFERENCES

1. J. H. Winters, “On the Capacity of Radio Communication Systems with Diversity in a Rayleigh Fading Environment,” *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5 (June 1987): 871–877.
2. G. J. Foschini and M. J. Gans, “On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas,” *Wireless Personal Communications*, vol. 6 (March 1998): 311–335.
3. M. Hata, “Empirical Formula for Propagation Loss in Land Mobile Services,” *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3 (August 1980): 317–325.
4. T. Zwick, C. Fischer, and W. Wiesbeck, “A Stochastic Multipath Channel Model Including Path Directions for Indoor Environments,” *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6 (August 2002): 1178–1192.
5. C. Waldschmidt, T. Fügen, and W. Wiesbeck, “Spiral and Dipole Antennas for Indoor MIMO Systems,” *IEEE Antennas and Wireless Propagation Letters*, vol. 1, no. 9 (2002): 176–178.
6. T. Fügen, J. Maurer, T. Kayser, and W. Wiesbeck, “Capability of 3D Ray Tracing for Defining Parameter Sets for the Specification of Future Mobile Communications Systems,” *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11 (November 2006): 3125–3137.
7. J. Maurer, T. Fügen, and W. Wiesbeck, “A Ray-Optical Channel Model for Vehicle-to-Vehicle Communications,” *Proceedings in Physics: Fields, Networks, Computational Methods, and Systems in Modern Electrodynamics*, P. Russer, M. Mongiardo (eds.) (December 2004): 243–254.
8. J. Maurer, C. Waldschmidt, T. Kayser, and W. Wiesbeck, “Characterization of the Time-Dependent Urban MIMO Channel in FDD Communication Systems,” *Proc. IEEE Vehicular Technology Conference VTC-Spring* (April 2003): 544–548.
9. T. Fügen, J. Maurer, C. Kuhnert, and W. Wiesbeck, “A Modeling Approach for Multiuser MIMO Systems Including Spatially-Colored Interference,” *Proceedings IEEE Global Telecommunications Conference* (December 2004).
10. J. R. Pierce and S. Stein, “Multiple Diversity with Nonindependent Fading,” *Proc. Institute of Radio Engineers (IRE)*, vol. 48 (January 1960): 89–104.

11. J. Bach Andersen and F. Hansen, "Antennas for VHF/UHF Personal Radio: a Theoretical and Experimental Study of Characteristics and Performance," *IEEE Transactions on Vehicular Technology*, vol. 26, no. 4 (November 1977): 349–357.
12. C. Waldschmidt, S. Schulteis, and W. Wiesbeck, "Complete RF System Model for the Analysis of Compact MIMO Arrays," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3 (May 2004): 579–586.
13. P.-S. Kildal and K. Rosengren, "Correlation and Capacity of MIMO Systems and Mutual Coupling, Radiation Efficiency, and Diversity Gain of Their Antennas: Simulations and Measurements in a Reverberation Chamber," *IEEE Communications Magazine*, vol. 42, no. 12 (December 2004): 104–112.
14. V. Jungnickel, V. Pohl, and C. von Helmolt, "Capacity of MIMO Systems with Closely Spaced Antennas," *IEEE Communications Letters*, vol. 7, no. 8 (August 2003): 361–363.
15. C. Waldschmidt, C. Kuhnert, M. Pauli, and W. Wiesbeck, "Handy MIMO," *IEE Communications Engineer* (February/March 2005): 22–25.
16. S. Schulteis, C. Kuhnert, and W. Wiesbeck, "Performance of a PDA Equipped with Three Dual-Band Inverted-F antennas for MIMO and Diversity Systems," *Proceedings Int. ITG/IEEE Workshop on Smart Antennas* (March 2006).
17. G. Fettweis, M. Löhning, D. Petrovic, et al, "Dirty RF: A New Paradigm," *Proceedings IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications* (September 2005).

Chapter 58

Multipath Techniques
for Handset/Terminal
Antennas

Per-Simon Kildal

Chalmers University of Technology

Charlie Orlenius

Bluetest AB

CONTENTS

58.1	INTRODUCTION.....	58-2
58.2	MULTIPATH ENVIRONMENT WITHOUT LINE-OF-SIGHT (LOS).....	58-2
58.3	CHARACTERIZATION OF SINGLE-PORT ANTENNAS IN MULTIPATH.....	58-4
58.4	CHARACTERIZATION OF MULTI-PORT ANTENNAS IN MULTIPATH.....	58-7
58.5	CHARACTERIZATION OF DIVERSITY AND MIMO PERFORMANCE	58-12
58.6	EMULATION OF MULTIPATH AND FADING USING REVERBERATION CHAMBER.....	58-18
58.7	MEASURING ANTENNAS AND ACTIVE TERMINALS IN REVERBERATION CHAMBER	58-25

58.1 INTRODUCTION

Mobile wireless terminals, such as handsets or phones, are subject to strong fading due to multipath propagation, particularly when used in urban and indoor environments. The performance of the terminals in such environments can be significantly improved by making use of antenna diversity. This means that the signals on two antennas (with different positions, polarizations, or radiation patterns) are combined in such a way that there are shallower fading minima on the combined signal than on each of the two contributing signals. Thereby, the signal-to-noise ratio (SNR) in the fading dips will be better, so that the fading margins in the system link budget can be reduced. The increased SNR can also be used to increase the capacity of the communication channel, if the system allows this. If there are several antennas on both the transmitter and the receiver side, it is possible to establish several communication channels through the multipath environment, which allows for even larger capacity increase. Such systems are often referred to as multiple input multiple output (MIMO) systems. This chapter describes how to characterize both single-port and multipoint antennas for use in multipath and in particular for achieving diversity and MIMO performance.

The so-called reverberation chamber¹ has, for a couple of decades, been used for the EMC testing of radiated emissions and immunity. This chamber is a metal cavity that is sufficiently large to support many resonant modes, and it is provided with means to stir the modes so that statistical field variations appear. It has been shown that the reverberation chamber represents a multipath environment of a similar type to what is found in urban and indoor environments. Therefore, in the last six years it has been extended to characterize antennas and terminals designed for use in environments with multipath propagation. This chapter will therefore also describe how reverberation chambers work and how they can be used to measure performance of antennas and active terminals, in particular when used with diversity and MIMO capability.

58.2 MULTIPATH ENVIRONMENT WITHOUT LINE-OF-SIGHT (LOS)

Traditionally, antennas were designed for use in environments where there is line-of-sight (LOS) between the two sides of the communication link. Therefore, antennas were characterized in the equivalent of a free space environment, by measurements in anechoic chambers. Still, there may in reality be additional wave paths between the transmitting and receiving sides. For example, large smooth objects located outside the LOS will cause reflection, edges of large objects will cause diffraction, and small or irregular objects will cause scattering of the waves originating from the transmitting antenna. The wave contributions via these paths will add at the receiving side. They have independent complex amplitudes (i.e., amplitudes and phases) that may add up constructively or destructively, or to anything between these two extremes. The wave paths and their complex amplitudes will also change fast with time, due to the moving of the terminal or parts of the environment, and thereby signal variations will be present at the receiver as a function of time, referred to as fading. The largest signal variations appear when the LOS is blocked. The previous type of fading is more accurately referred to as *small scale fading*, to distinguish it from the signal variations appearing only due to variations in distance from the transmitting antenna or due to partly shadowing and similar of the LOS, referred to as *large-scale fading*.²

Rayleigh Fading

The multipath environment can, at the receiving side, be characterized by several independent incoming plane waves. This independency means that their amplitudes, phases,

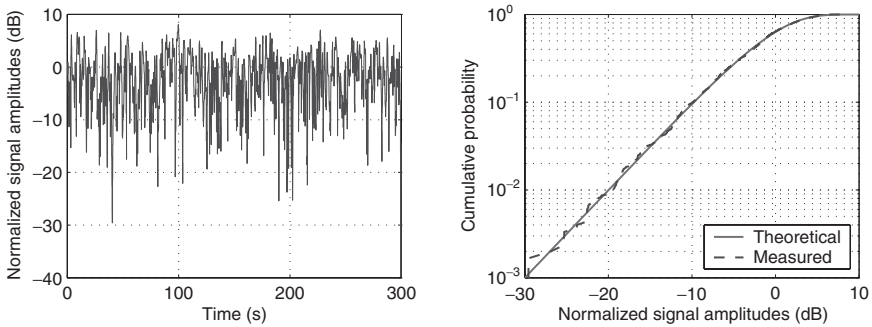


FIGURE 58-1 Example of a fading signal (left) and its cumulative probability distribution (CDF) function (right). The signal levels are presented in dB after being normalized to the time averaged power.

and polarizations as well as the angles of arrival (AoA) are arbitrary relative to each other. If the LOS is absent, and if the number of incoming waves is large enough (typically a few hundred), or if the antenna is moved around in a less rich environment, the in-phase and quadrature components of the received complex signal become normally distributed (Gaussian). From this, their associated magnitudes get a Rayleigh distribution, the power an exponential distribution, and the phases a uniform distribution over 2π . This is a direct result of the central limit theorem. Figure 58-1 shows an example of a Rayleigh faded signal. The signal varies by more than 25 dB. The Rayleigh fading is most conveniently illustrated by a cumulative probability density function (CDF), showing the cumulative probability of signal amplitudes in dB, also shown in Figure 58-1. The reference level for the dB values is in the figure taken to be the average received power, i.e., the total received power divided by the number of samples. There is a probability of 1‰ of having deeper dips than -30 dB, and of 1% of having deeper dips than -20 dB.

Angle of Arrival (AoA) Distribution, XPD and Polarization Imbalance

The arriving waves may have a certain AoA distribution in the elevation and azimuth planes. It is natural to assume that the mobile terminal can be oriented arbitrarily, relative to directions in the horizontal plane, which means that the azimuth angle is uniformly distributed. The terminals may, under normal use, have a certain preferred or most probable orientation relative to the vertical axis, and common environments (in particular outdoor) have larger probability of waves coming in from close-to-horizontal directions than close-to-vertical. Therefore, there may be a need for an elevation distribution function to describe real multipath environments. Real environments normally also have a larger content of vertical polarization than horizontal, because most base stations are vertically polarized. This is characterized, in propagation literature, by a cross-polar power discrimination (XPD).³ Instead, this chapter refers to XPD as a polarization imbalance in the environment, because it represents a deterioration of the isotropic reference environment defined in the next subsection. Both the AoA distribution and XPD are different in different real environments. This is problematic because the performance of the antennas and wireless terminals then depend on where they are used or measured, so that results of measurements in one environment cannot directly be transferred to another. Therefore, it is convenient to define a reference environment, which can easily be reproduced in practice, and which gives a repeatable result for the performance. The isotropic environment with polarization balance and a uniform AoA distribution over the whole sphere has this characteristic.

Isotropic Reference Environment

As explained previously, it is desirable to have an isotropic reference environment with polarization balance and a uniform distribution of AoA in both azimuth and elevation, i.e., an environment in which all polarizations and AoAs over the whole unit sphere are equally probable. This simplifies the characterization of antennas and terminals in the sense that the performance becomes independent of the orientation of the antenna in the environment. The reverberation chamber emulates such an isotropic environment. This reference environment has no true counterpart in reality, but it is still quite representative as any environment will appear isotropic if the terminal is used with arbitrary orientation in the environment. It is important here to point out that the antenna normally is not located symmetrically on the phone. Therefore, the phone has different orientations in the environment when used on the left and right sides of the head, as opposite sidewalls of the phone will point upwards for these two talk positions.

58.3 CHARACTERIZATION OF SINGLE-PORT ANTENNAS IN MULTIPATH

This section will describe how to characterize single-port antennas for use in multipath with Rayleigh fading. Naturally, this is strongly related to characterization in LOS.

Antenna Impedance, Port Impedance, and Reflection Coefficient

All antennas are characterized by their far field function $\mathbf{G}(\theta, \varphi)$ and input impedance Z_a . The far field function can be used to plot the co- and cross-polar radiation patterns and to find traditional LOS quantities such as directivity and sidelobe levels. For antennas operating in an isotropic multipath environment, the shape of the far field function does not play any significant role, particularly not for single-port antenna systems. In multipoint antennas it has some significance as it determines the correlation between received signals on the ports. Therefore, we will need to define $\mathbf{G}(\theta, \varphi)$ as well. In order to connect to other components like receivers and transmitters, the antenna must have a well-defined port (or ports), often in the form of a transmission line or a connector for a transmission line. The characteristic impedance Z_c of this transmission line then defines the port impedance, normally 50 Ohms. When the port impedance is known, the antenna impedance Z_a can be transformed to a complex reflection coefficient r on the transmission line, where

$$r = (Z_a - Z_c) / (Z_a + Z_c) \quad (58-1)$$

In an S -parameter representation of the antenna, $r = S_{11}$.

Definition of Far Field Function

If the far field function is mathematically formulated in transmit mode (caused by a given source current at the antenna port) as a complex vector field function, it can be included in the equivalent circuit of the same antenna in receive mode, as given in the next subsection. This complete complex radiation field function (containing even phase information) is also needed in theoretical multipath characterization in order to add the received voltages (with correct amplitudes and phases) from several simultaneous incident waves with different AoAs.

The far field of any antenna will have the form

$$\mathbf{E}(r, \theta, \varphi) = \frac{1}{r} e^{-jk_r r} \mathbf{G}(\theta, \varphi) \quad (58-2)$$

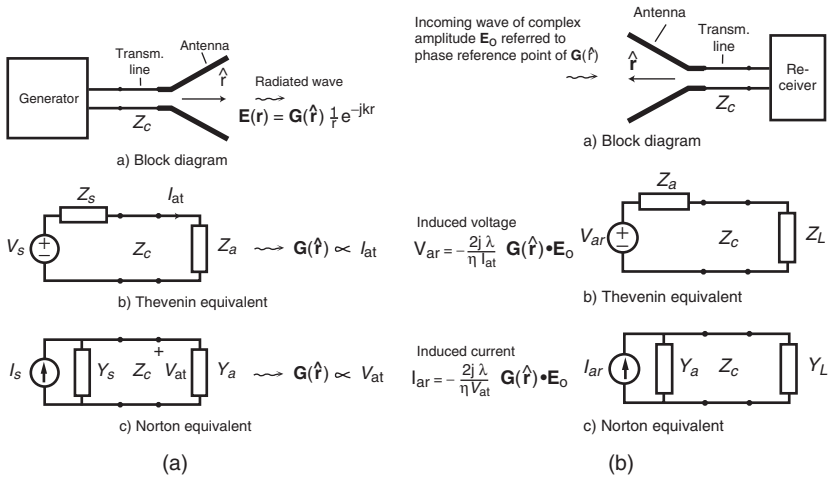


FIGURE 58-2 Transmit (left) and receive (right) equivalent circuits of antenna with far field function $\mathbf{G}(\hat{\mathbf{r}})$, input impedance $Z_c = 1/Y_c$, and port impedance Z_c . The receive case is from Section 2.5.5.⁴ The parameters in the formulas are explained in the text.

where r is the distance from the phase reference point (center of coordinate system) to the observation point, and $\mathbf{G}(\theta, \varphi)$ is the far field function.⁴ $\mathbf{G}(\theta, \varphi)$ is a complex vector function of direction (θ, φ) from the phase reference point to the observation point. It is written alternatively as $\mathbf{G}(\hat{\mathbf{r}})$, where $\hat{\mathbf{r}}$ is a unit vector in the direction (θ, φ) . The far field function is shown in the equations in Figure 58-2. The total radiated power expressed by $\mathbf{G}(\theta, \varphi)$ becomes

$$P_{\text{rad}} = \frac{1}{2\eta} \iint_{\Omega} |\mathbf{G}(\theta, \varphi)|^2 d\Omega \quad (58-3)$$

with η the free space wave impedance. The $1/2$ factor comes because of the use of peak-valued complex vectors rather than effective-valued ones.

The far field function can be determined analytically, by measurements, or by simulations. In the latter case in order to avoid mistakes, it is safest to use a realizable source at the port, i.e., either of the two generator models (Thevenin or Norton) in the left part of Figure 58-2. The source impedance Z_s must be chosen to be equal to the defined port impedance Z_c . Often, the far field $\mathbf{E}(\mathbf{r}) = \mathbf{E}(r, \theta, \varphi)$ is computed at chosen directions (r, θ, φ) on a sphere of fixed radius r around the antenna. The far field function is then determined from the far field in these points by

$$\mathbf{G}(\theta, \varphi) = \mathbf{E}(r, \theta, \varphi) r e^{jkr} \quad (58-4)$$

provided r is large enough. The far field function in receive mode is the same as in transmit mode.

Equivalent Circuits of Transmit and Receive Antennas

The equivalent circuit of an antenna is well known when it is transmitting, as shown in the left part of Figure 58-2. But, when there is a need to analyze diversity and MIMO

antennas theoretically, the equivalent circuit in receive mode is needed. Receive equivalent circuits are given in many textbooks, but they are often not complete. The formulas for induced open-circuit port voltage V_{ar} in the Thevenin equivalent and the induced short-circuit port current I_{ar} in the Norton equivalent can be given explicitly from the far field function and the incident wave E_0 . Such complete equivalent circuits can be found though,⁴ and they are also included in the right part of Figure 58-2. The formula for V_{ar} (alternatively I_{ar}) is given in terms of the following parameters: the wavelength λ , the free space wave impedance η , the far field function $G(\theta, \varphi)$ calculated for the transmit mode as explained previously, and finally by the current I_{at} into the antenna port (alternatively the voltage V_{at} at the antenna port) when the far field function $G(\theta, \varphi)$ of the antenna was calculated. See the definitions I_{at} (alternatively V_{at}) in the left part of Figure 58-2.

Far field functions should, as explained previously, always be calculated when the antenna is excited by a complete Thevenin or Norton equivalent circuit for an impedance-matched generator at the input, i.e., using $Z_s = Z_c$ or $Y_s = Y_c = 1/Z_c$. Then,

$$I_{at} = V_s / (Z_c + Z_a) \quad (58-5)$$

in the formula for V_{ar} in the Thevenin equivalent circuit for the receiving case, and

$$V_{at} = I_s / (Y_a + Y_c) = I_s Z_a Z_c / (Z_a + Z_c) \quad (58-6)$$

in the formula for I_{ar} in the Norton equivalent circuit for the receiving case.

Mean Effective Gain (MEG) and Mean Effective Directivity (MED)

Antennas in fading environments are sometimes characterized by the so-called mean effective gain (*MEG*). The formula for *MEG*^{3,5} can be calculated from the far field function of the antenna, and it is a function of the orientation of the antenna, its polarization, and the azimuth and elevation distributions of the AoA in the environment, as well as the XPD of the environment. For the isotropic environment, the *MEG* becomes equal to half the classical total radiation efficiency e_{rad} , i.e.,

$$MEG = e_{rad} / 2 \quad (58-7)$$

Actually, for an arbitrary environment the *MEG* can be separated in two factors: the classical total radiation efficiency and a mean effective directivity (*MED*),⁶ i.e.,

$$MEG = MED \cdot e_{rad} \quad (58-8)$$

In this way the *MED* solely contains the effect of the environment and the shape of the radiation pattern, whereas the radiation efficiency contains the effects of losses and impedance mismatch. The formula for *MED* is the same as for *MEG*, but the realized gain for the far field functions in the formula for *MEG* must be replaced by directive gain far field functions in the formula for *MED*, where the realized and directive gains are standard IEEE definitions. Thereby, the relationship between *MEG* and *MED* becomes the same as the relationship between the classical realized and directive gains for LOS systems (radiation efficiency). The radiation efficiency can, for instance, be measured in a reverberation chamber. The formulas for the *MEG* and *MED* are not shown here, because the consideration here is isotropic environments.

Total Radiation Efficiency

From the previous subsections, it can be stated that for antennas in an isotropic multipath environment the performance is determined only by the total radiation efficiency and not by the shape of the radiation pattern or the realized gain. The definition for the total radiation efficiency used here is the ratio between the total radiated power and the maximum available power that can be delivered from the generator. Thus,

$$e_{\text{rad}} = P_{\text{rad}} / P_{\text{in,max}} \quad (58-9)$$

where the maximum available power appears when $Z_a = Z_c$ and becomes either

$$P_{\text{in,max}} = \frac{1}{2} Z_c |I_{\text{at}}|^2 \quad \text{or} \quad P_{\text{in,max}} = \frac{1}{2} |V_{\text{at}}|^2 / Z_c$$

with I_{at} and V_{at} defined previously, depending on whether the Thevenin or Norton equivalents are used, respectively. For a lossless antenna, this is known to be

$$e_{\text{rad}} = 1 - |r|^2 \quad (58-10)$$

where r is the complex reflection coefficient at the antenna port defined in Eq. 58-1.

58.4 CHARACTERIZATION OF MULTIPOINT ANTENNAS IN MULTIPATH

In LOS systems it is possible to increase the signal-to-noise ratio by increasing the directivity of the antennas, because the received signal increases with the gain of both the transmit and receive antennas. However, in the previous section it was shown that antennas in isotropic multipath environments do not get improved performance by increasing directivity. Still, several antenna elements can be used to form a diversity or MIMO system, but to get improved performance the signals must be processed in a more advanced way than in classical arrays.

To realize antenna diversity and MIMO systems, several antenna elements need to be located together in an array. This is not a classical array, in which the antennas are combined with certain amplitude and phase relations between them. Therefore, the diversity or MIMO antenna is called a multiport antenna rather than an array, meaning that each port can be characterized as a single antenna, and that the signal inserted (transmitting case) or being present (receiving case) at each port has no prescribed amplitude and phase relation to the signals on the other ports. Instead, the signals are processed incoherently to optimize system performance.

Definition of Channel When Using Multiport Antenna Systems in Multipath

The propagation channel is defined with the antennas included. Therefore, a channel starts at one transmit port connected with one element or beam of the transmit antenna and ends up at a receive port connected with one element or beam of the receive antenna. Thus, if the transmit antenna has M ports and the receive antenna has N ports, $M \times N$ channels can be generated through the environment, as illustrated in Figure 58-3.

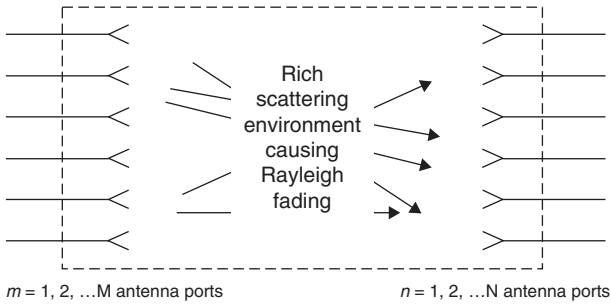


FIGURE 58-3 Definition of MIMO system with multiport antennas on both the transmitting and receiving sides. There are in total $M \times N$ channels through the environment, each one defined between one input port of the transmit antenna and one output port of the receive antenna.

Isolated Elements

Classical single-beam array antennas for LOS systems have one port and are characterized by their far field functions and antenna impedances, in the same way as other single-port antennas. The latter is the impedance seen at the array port when all elements are excited with the amplitude and phase that give the desired shape of the far field function. The far field function of a classical array antenna is the product of two factors: the far field function of an isolated element (which is equal for equal elements at different location) and the array factor. The far field function of one isolated element is the far field function of this element when all the other elements are removed, but with the ground plane on which they are located present.

In theoretical work on diversity and MIMO systems, the element antennas are sometimes treated as if they were isolated, even though an array factor cannot be defined because the signals at every element are processed independently. However, this treatment must be done with care because the isolated element approach is strictly not valid when analyzing multiport antennas. It will only give correct results if the ports are completely uncoupled, which they rarely will be in reality, except for two-port antennas with orthogonal polarizations on the two ports. The correct analysis is done by using the embedded element approach as described next.

Embedded Elements

The embedded element approach also has its background in classical array analysis. However, because it is regarded as an advanced topic compared to the isolated element approach, it is rarely covered in introductory antenna courses. The embedded element approach considers the far field of each single array element when all the other elements are present, but they are not excited and instead terminated with their port impedances. The far fields of such embedded elements are sometimes called *active element patterns*, a terminology introduced previously,⁷ but they are now more commonly and descriptively referred to as *embedded element patterns*.^{6,8} The excited antenna will induce radiating currents on the terminated non-excited antennas. Therefore, the far field function of the embedded element may be very different from that of the isolated element, in particular if the elements are closely spaced with large mutual coupling. The far field functions of the embedded elements are used to describe scan blindness in classical arrays, whereas in MIMO and diversity antennas they

play an even more significant role. In a multipath environment the received signal on each port is transmitted or detected independently relative to the signals on the other ports, and therefore each port transmits or receives signals through their embedded far field functions. In such a way, the radiation efficiency at each port as well as the correlation between the signals at all ports, are determined by the far field functions of the corresponding embedded elements. Both these quantities (radiation efficiency and correlation) are needed in order to quantify performance of diversity and MIMO antenna systems. Such analysis in terms of far field functions of embedded elements was introduced^{6,8} based on the analysis of a diversity antenna consisting of two parallel dipoles.⁹ The far field functions of embedded elements can be computed by most commercial computer programs, by exciting the port to be computed and terminating all non-excited ports with their port impedances (normally 50 Ohm). The embedded element patterns can also be measured in anechoic chambers. However, it is also possible to characterize diversity and MIMO antennas without knowing the embedded element patterns explicitly. For example, the reverberation chamber provides a way of measuring radiation efficiency and correlation without going via the embedded element patterns. It can also be used to provide measured estimates of complete communication channels, and from these the diversity gain and maximum available capacity of complete diversity and MIMO antenna systems can be obtained.

Embedded Radiation Efficiency

This subsection starts with an explanation of how to calculate the total radiation efficiency of an embedded element of a multiport antenna. Figure 58-4 illustrates a diversity antenna consisting of two dipoles and with receivers and/or transmitters on both ports. This means that the diversity combination is implemented after the receivers, and that the diversity separation is implemented in front of the transmitters. This configuration was analyzed¹⁰ by using a classical semi-analytical approach represented by the equivalent circuit in the upper part of Figure 58-5. The equivalent circuit is generally applicable to any two-port antenna consisting of single-mode radiating elements, such as dipoles, slots, and—to some degree—patches. The most accurate way to calculate this is to perform a complete numerical computation using a full-wave code of any kind. This way there is no restriction to the elements being of single-mode type. The general form of the equivalent circuit is shown in the lower part of Figure 58-5. This latter equivalent circuit doesn't show the induced currents, so it cannot be used to calculate or understand why the far field function is different from that of an isolated antenna element. This general simplified equivalent circuit is simply an illustration of the result of a numerical calculation of Z_a and $\mathbf{G}(\theta, \varphi)$.

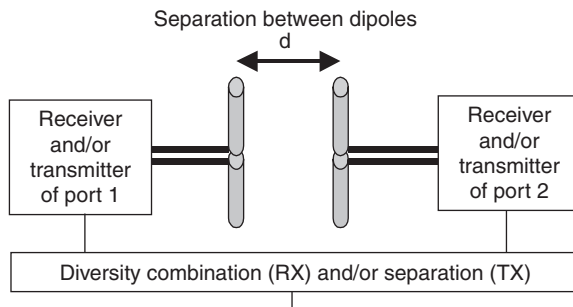


FIGURE 58-4 Illustration of diversity antenna consisting of two parallel dipoles. There are receivers and/or transmitters at each of the two dipole ports.

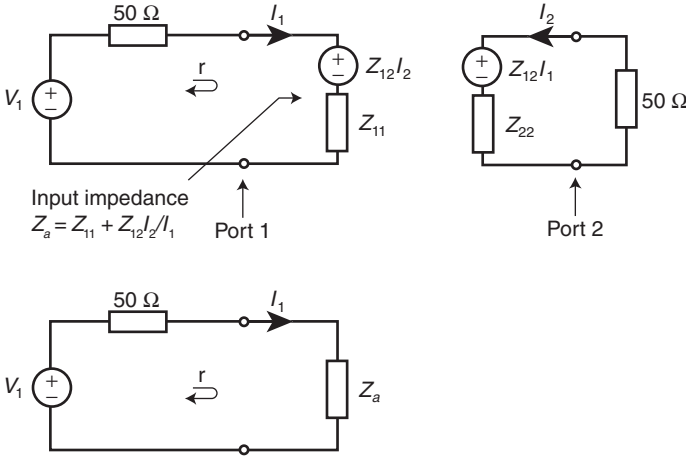


FIGURE 58-5 Upper figure: Complete equivalent circuit of a diversity antenna consisting of two single-mode elements with ports 1 and 2, respectively, when port 1 is excited and port 2 terminated with the port impedance (50 Ohms), for calculation of the far field function and radiation efficiency of the embedded element 1. Lower figure: The same equivalent circuit on the general Thevenin form in Figure 58-2b, left case. The port impedance is $Z_c = 50$ Ohm.

If the diversity antenna itself consists of materials with very small or no losses, the only contributions to the total radiation efficiency of the embedded element will be the absorption in the termination on port 2, and the reflections on the excited port 1. Using the upper equivalent circuit in Figure 58-5, the input impedance of the embedded element at port 2 is

$$Z_a = Z_{11} + \frac{Z_{12}I_2}{I_1} = Z_{11} - \frac{Z_{12}^2}{Z_{22} + Z_c} \quad (58-11)$$

where the rightmost expression is obtained by using the equivalent circuit of port 2 in Figure 58-5 (upper figure). Z_{11} and Z_{22} are the impedances of the two isolated element antennas and Z_{12} is their mutual impedance. Note that Z_a is the impedance when the equivalent circuit of port 1 has the general Thevenin form as shown in Figure 58-5 (lower figure). The power absorbed in the antenna impedance Z_a represents both the radiated power and the power absorbed in the 50 Ohms termination on port 2. The reflection coefficient r at port 1 and the corresponding mismatch efficiency e_{refl} become

$$r = \frac{Z_a - Z_c}{Z_a + Z_c} \quad \text{and} \quad e_{\text{refl}} = 1 - |r|^2 \quad (58-12)$$

The total power that is accepted by port 1 is found as $P_{\text{acc}} = \text{Re}\{Z_a\}|I_1|^2/2$. The difference between this and the power $P_2 = Z_c|I_2|^2/2$ absorbed in the load on port 2 is the radiated power P_{rad} , i.e. $P_{\text{rad}} = P_{\text{acc}} - P_2$. Finally, the total radiation efficiency becomes

$$e_{\text{rad}} = e_{\text{refl}}e_{\text{abs}} \quad \text{with} \quad e_{\text{abs}} = \frac{P_{\text{rad}}}{P_{\text{acc}}} = 1 - \frac{Z_c|I_2|^2}{\text{Re}\{Z_a\}|I_1|^2} \quad (58-13)$$

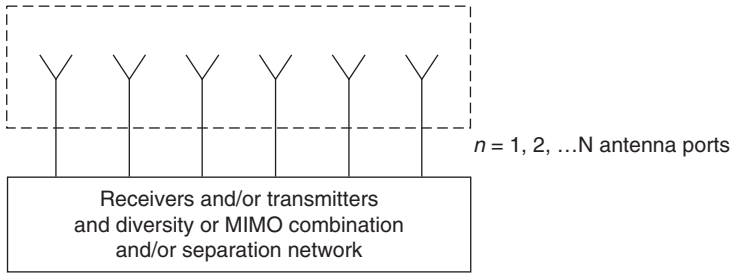


FIGURE 58-6 Illustration of N -port antenna that can be characterized by its S -parameters between all ports

where e_{abs} is the efficiency due to the power dissipated in the load on port 2. This can be expressed much simpler by using the S -parameters between ports 1 and 2, according to

$$e_{\text{rad}} = 1 - |S_{11}|^2 - |S_{12}|^2, \quad e_{\text{refl}} = 1 - |S_{11}|^2, \quad e_{\text{abs}} = \frac{1 - |S_{11}|^2 - |S_{12}|^2}{1 - |S_{11}|^2} \quad (58-14)$$

The S -parameter representation can readily be extended to any multiport antenna, such as that illustrated in Figure 58-6. The total embedded radiation efficiency of embedded element number i —when there are a total of N ports—becomes

$$e_{\text{rad}} = 1 - \sum_{j=1}^N |S_{ij}|^2 \quad (58-15)$$

There will be an additional efficiency factor if there are losses in the materials that the antenna is made from. The last equation defines a fundamental limitation also of elements in classical dense arrays.¹¹ If the elements are very close (typically closer than 0.5 wavelengths), the mutual couplings cause a severe efficiency reduction. In a classical array this fundamental limitation causes the gain not to increase proportional to the number of elements unless the element spacing is kept larger than 0.5 wavelengths.

Correlation Between Ports

The correlation coefficient ρ between received signals on two ports can be calculated from the correlation between the induced voltages in the receive equivalent circuits of the two ports, when these voltages are a result of multipath due to incident plane waves with uniform AoA distribution and polarization balance. This formula for evaluating the correlation by using a statistical distribution of induced voltages can readily be transformed to deterministic integrals in terms of the normalized coupling between the far field functions of the two ports:³

$$\rho = \frac{\iint_{4\pi} \mathbf{G}_1(\theta, \phi) \cdot \mathbf{G}_2^*(\theta, \phi) d\Omega}{\sqrt{\iint_{4\pi} \mathbf{G}_1(\theta, \phi) \cdot \mathbf{G}_1^*(\theta, \phi) d\Omega \iint_{4\pi} \mathbf{G}_2(\theta, \phi) \cdot \mathbf{G}_2^*(\theta, \phi) d\Omega}} \quad (58-16)$$

where $\mathbf{G}_1(\theta, \phi)$ and $\mathbf{G}_2(\theta, \phi)$ are the embedded far field functions of ports 1 and 2, respectively.

58.5 CHARACTERIZATION OF DIVERSITY AND MIMO PERFORMANCE

With diversity, two antennas are used, which are located sufficiently far from each other (space diversity) or with different polarizations (polarization diversity) or radiation patterns (pattern diversity). The received signal will then be uncorrelated on the two ports, and it is very unlikely that there will be a fading dip simultaneously on both ports. Therefore, by an appropriate combination of the two signals, the probability of a fading dip in the combined signal will be strongly reduced. There are several different possible combination schemes, such as switch diversity, selection combining, or maximum ratio combining, and the improvement of the CDF can be as large as 10 to 12 dB at 1% value of the CDF. The details of the different diversity schemes are not covered here.³ Selection combining will be assumed in the discussions to follow.

Diversity Gains (Apparent, Effective and Actual)

As an example, the two parallel dipoles in Figure 58-4 were characterized at 900 MHz and with a given separation between them. Figure 58-7 shows the cumulative distribution function (CDF) of fading signal amplitudes measured in a reverberation chamber. The curve of the reference antenna (a dipole with known radiation efficiency) follows the theoretical Rayleigh distribution very closely (because the curve is corrected for the known radiation efficiency of the reference antennas), which is the guarantee for a rich scattering environment. The CDF at each port of the diversity antenna also has the same shape as a theoretical Rayleigh distribution, but they are shifted to the left because their total radiation efficiency is lower. Actually, the horizontal spacing between the CDFs of the two ports of the diversity antenna and the CDF of the ideal reference is equal to the embedded radiation efficiency of the diversity antenna ports. If selection combining is applied to the signal samples on the two ports, an improved CDF is shown in the figure. The diversity gain is the difference between the selection-combined CDF and a reference CDF at a certain CDF-level, normally chosen to be 1%. It is possible to distinguish between apparent, effective, and actual diversity gains¹²⁻¹³ depending on the reference.

- **Apparent diversity gain** Difference between power levels in dB (at certain CDF level), between CDF of combined signal, and CDF of signal at the port with the strongest average signal levels.
- **Effective diversity gain** Difference between power levels in dB (at certain CDF level), between CDF of combined signal, and CDF of signal at the port of an ideal single antenna (corresponding to radiation efficiency of 100%), measured in the same environment.
- **Actual diversity gain** Difference between power levels in dB (at certain CDF level), between CDF of combined signal, and CDF of signal at the port of an existing practical single antenna that is to be replaced by the diversity antenna under test, measured at the same location (for example, relative to a head phantom).

An ideal single antenna means that the antenna has 100% total radiation efficiency, or the equivalent thereof, obtained by correcting the CDF of a practical reasonable well-matched antenna with its finite radiation efficiency. In the case of actual diversity gain, the practical antenna reference shall be located in the position relative to an object (for example a head phantom) that corresponds to the desired position of operation for the existing antenna as well as the replacing diversity antenna. Under these definitions, the actual diversity gain in dB is the effective diversity gain in dB minus the radiation efficiency in dB of the single existing

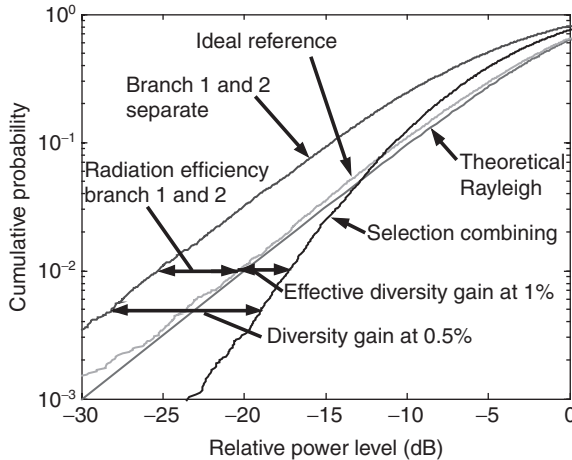


FIGURE 58-7 Cumulative probability distribution function (CDF) of measured transmission function in reverberation chamber for ideal reference antenna (corrected for its finite radiation efficiency) and diversity antenna consisting of two parallel dipoles⁶

antenna that the diversity antenna shall replace. The effective diversity gain represents the gain over a single ideal reference antenna, where the latter is measured with no additional antenna close to it.

As shown in Figure 58-7 the apparent diversity gain at 1% CDF level is 8 dB, whereas the effective diversity gain compared to the ideal single antenna reference is only about 3 dB. This means that in a system specified for a fading margin of 20 dB to be able to receive with sufficient quality 99% of the time (for example, 1% CDF level), the fading margin can be reduced by 3 dB if this specific diversity antenna (two parallel dipoles with 15 mm spacing) is used instead of a very good single antenna. The diversity gain could easily be made larger by using larger dipole spacing, or by choosing two orthogonal antennas. The theoretical maximum is 10 dB by selection combining. The discrepancy between 3 dB and 10 dB is, in the example, mainly due to the low radiation efficiency when the dipoles are spaced 15 mm apart. This is caused by mutual coupling, giving large absorption in the 50 Ohm load of the not-excited dipole. The reduced diversity gain due to correlation has, in comparison, a small effect, as explained next.

There are published results of diversity gains as a function of dipole spacing.⁶ There is also significant actual diversity gain for antennas used near the human body (a simple cylindrical head phantom). The actual diversity gains at 900 MHz is shown to be more than 6 dB, even when the dipole separation is only 2 cm. This is very promising for the use of diversity in mobile phones. Larger diversity gains will naturally be present for antennas with uncoupled orthogonally polarized ports.

Theoretical Determination of Diversity Gain

Theoretically, the effective diversity gain G_{eff} can be determined by the following formula:

$$G_{\text{eff}} = e_{\text{rad}} G_{\text{app}} \quad (58-17)$$

where G_{app} is the apparent diversity gain. This depends on the correlation coefficient ρ obtained from Eq. 58-16 by using the far field functions of the two embedded ports. The relation is approximately given by the following formula.¹⁴

$$G_{app} = 10 \cdot e_\rho \quad \text{with} \quad e_\rho = \sqrt{1 - |\rho|^2} \quad (58-18)$$

where 10 is the maximum apparent diversity gain at 1% CDF level with selection combining, and e_ρ is an approximate expression for the correlation efficiency, i.e., the reduction in diversity gain due to correlation between the signals on the two ports. This formula is not very accurate for correlations close to unity when compared with the correct formulas.¹⁴ However, if ρ is scaled with a factor 0.99, the formula becomes

$$e_\rho = \sqrt{1 - |0.99 \cdot \rho|^2} \quad (58-19)$$

which differs from the correct expression for the apparent diversity gain at 1% CDF by less than 0.1 dB (as shown in Figure 58-8). According to the graph, the correlation must be very strong in order to get significant reduction of the diversity gain.

Maximum Available Spectral Capacity of MIMO System from Shannon

Future mobile communication systems propose to make use of multiport antennas on both the base station and terminal sides to form several communication channels, i.e., a MIMO system. For example, six antennas on both sides form $6 \times 6 = 36$ possible communication channels. The data is then on transmit distributed among the channels and combined again after reception, in such a way that the overall channel capacity is maximized. This means that those of the 36 channels that provide fading maxima transfer much more bits per second than those having fading minima.

The capacity of a communication link is measured in terms of the bit rate, i.e., the number of bits per second (bits/sec). When comparing different types of communication

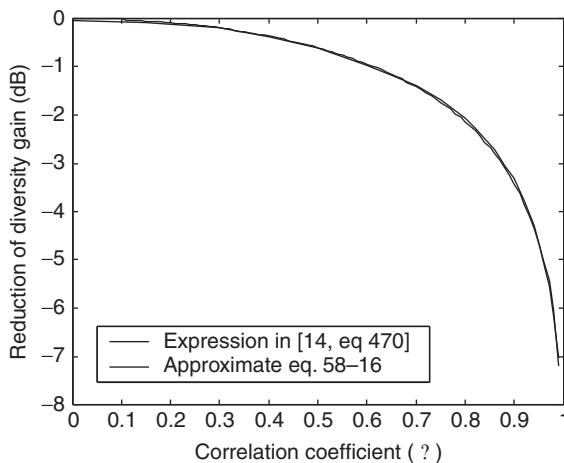


FIGURE 58-8 Correlation efficiency at 1% CDF, i.e., reduction of diversity gain at 1% cumulative probability level, as a function of the correlation ρ in Eq. 58-16, obtained with the approximate Eq. 58-19 and the correct expression¹⁴

systems and modulation methods, it is also important to know how efficiently they use the allocated frequency spectrum. This can be quantified in terms of a spectral capacity (bits/sec/Hz), i.e., how many bits per second that can be transferred per frequency unit Hertz. The spectral capacity will naturally increase with SNR , which therefore must be specified or known. The spectral capacity (bits/sec/Hz) is in communication engineering sometimes also called the spectral efficiency, but in this text we will use the former expression to avoid confusion with antenna efficiencies.

The maximum possible average spectral channel capacity in a MIMO system can be calculated by using a version of Shannon's fundamental formula for capacity that is extended to multiport systems. This formula and the results for a 1×6 and a 3×6 MIMO system measured in a reverberation chamber will be discussed next.

Single-Port System

The maximum available spectral capacity of a communication channel through any environment is given by Shannon's classical formula

$$C = \log_{(2)} |1 + SNR| \quad (\text{bits / sec / Hz}) \quad (58-20)$$

where SNR is the signal-to-noise ratio at the receive side, and where $\log_{(2)}$ means the logarithm with base 2. This means that the capacity of a system does not only depend on its bandwidth, but also on the SNR . Thus, if the transmit power is increased, it is possible to transfer more bits/sec/Hz over the channel. However, in order to benefit from this we have to know the SNR on the channel. Therefore, estimation of the SNR is important in communication engineering, as is the feedback of this estimate to the transmit side for controlling the capacity. The Shannon formula is also valid under fading, but then the SNR of the channel varies with time, in which case channel estimation must be performed continuously and thereby becomes a very important topic.

Parallel Channels in LOS

If two parallel communication systems are used instead of one, the capacity can be naturally doubled if they are mounted in such a way that there is no interference. This can be done by locating the antennas of the extra system in the nulls of the radiation patterns of the other. Thus, the maximum spectral capacity of two such systems corresponding to two parallel channels becomes

$$C = C_1 + C_2 = \log_{(2)} \left| 1 + \frac{1}{2} SNR_1 \right| + \log_{(2)} \left| 1 + \frac{1}{2} SNR_2 \right| = \log_{(2)} \left(\left| 1 + \frac{1}{2} SNR_1 \right| \left| 1 + \frac{1}{2} SNR_2 \right| \right) \quad (58-21)$$

where the indexes on C and SNR denote systems 1 and 2. The factor $\frac{1}{2}$ on all the SNR s means that the total power has been divided equally between the two channels. Next is a discussion of whether or not it is possible to increase the capacity by combining the two antennas on each side to a classical directive array so that a single communication system with larger SNR is formed.

The signal to noise ratio of a LOS system is given by

$$SNR = \frac{P_r}{N} = \frac{P_t G_t G_r \left(\frac{\lambda}{4\pi r} \right)^2}{N} \quad (58-22)$$

where P_r is the received power, P_t is the transmitted power, G_r is the realized gain of the receive antenna, G_t is the realized gain of the transmit antenna, N is the noise power, and $\lambda/4\pi r$ is the free space attenuation. The gains on both the transmit and receive sides are now doubled by combining on each side two equal uncoupled directive antennas to a classical directive array with the double gain. Then, the SNR of the new system with classical two-element arrays on both sides is

$$SNR_{2 \times 2 \text{ arrays}} = 4SNR \quad (58-23)$$

or four times larger than the SNR of each of the two systems alone. The condition for this equation to be valid is that the noise power is independent of antenna gain.

The spectral capacity for the two cases, two parallel systems (which we may call two 1×1 systems), and the array case (for example, the 2×2 case) become (when the power is equally distributed between the two transmit antennas)

$$C_{\text{two } 1 \times 1} = 2 \log_{(2)} \left| 1 + \frac{1}{2} SNR \right| \xrightarrow{SNR \text{ large}} -2 + 2 \log_{(2)} |SNR| \quad (\text{bits / sec / Hz}) \quad (58-24)$$

$$C_{2 \times 2 \text{ array}} = \log_{(2)} |1 + 4SNR| \xrightarrow{SNR \text{ large}} 2 + \log_{(2)} |SNR|$$

These two equations can be plotted and compared to show that as long as $SNR > 11$ dB in each channel it is advantageous (for maximizing capacity) to use parallel systems with low gain antennas rather than one system with high gain antennas, and the maximum available spectral capacity increases by up to a factor 2 when the number of channels (for large SNR) is doubled. These are theoretical results. In practice there will be other limitations that may change this conclusion, such as hardware and cost constraints.

Parallel Channels in Rayleigh Fading

In a fading environment, the instantaneous spectral capacity is given by the same formulas as for the LOS case, but the formula for the SNR is different. The space attenuation is much larger due to the scattering from all blocking objects, and the SNR will fade with the fading signal level. Therefore, it is common to use an average maximum spectral capacity in characterization of systems for multipath, obtained by averaging the Shannon capacity over the statistical distribution of the power samples of the signal. The power samples are distributed exponentially in Rayleigh fading.

In a MIMO system with multiple ports on both the transmitting and receiving sides, there are many communication channels through the environment. Now, the signal (from one of the two transmit antenna ports in the example in the previous section) will be received on both the receive antenna ports. However, the signals on the two receive ports will be more or less uncorrelated (even if they are the same in terms of system modulation) due to the statistics of the fading. If it is assumed that they are uncorrelated, the maximum instantaneous spectral capacities for the similar cases as in the LOS example become

$$\tilde{C}_{1 \times 1} = \log_{(2)} (1 + \tilde{SNR}_{t1r1})$$

$$\tilde{C}_{1 \times 2 \text{ MIMO}} = \log_{(2)} (1 + \tilde{SNR}_{t1r1} + \tilde{SNR}_{t1r2})$$

$$\tilde{C}_{2 \times 2 \text{ MIMO}} = \log_{(2)} \left(\left(1 + \frac{1}{2} \tilde{SNR}_{t1r1} + \frac{1}{2} \tilde{SNR}_{t1r2} \right) \left(1 + \frac{1}{2} \tilde{SNR}_{t2r1} + \frac{1}{2} \tilde{SNR}_{t2r2} \right) - \text{channel coupling terms} \right) \quad (58-25)$$

The \tilde{C} and \tilde{SNR} notation with the tilde symbol means that C and SNR fade with time. The \tilde{C} 's need to be averaged over the distribution function of the signal level in order to compare the channels. It can be shown that the largest average capacity will be obtained if the different channels fade independently, so that the correlations between their received signals are low. Still, by comparing the spectral capacity formulas it can be seen that a MIMO array in a multipath environment has corresponding spectral capacity increase as parallel channels in a LOS environment, when compared to using single channel systems in the same two environments, respectively. The general formula for the maximum available capacity of a MIMO system with a transmit antenna with M ports and a receive antenna with N ports is

$$\tilde{C} = \log_2 \left(\det \left(\mathbf{I}_{NxM} + \frac{SNR}{M} \tilde{\mathbf{H}}_{NxM} \tilde{\mathbf{H}}_{NxM}^* \right) \right) \quad (58-26)$$

where \mathbf{I}_{NxM} is a unit matrix, $\tilde{\mathbf{H}}_{NxM}$ is a normalized complex channel matrix, and $\tilde{\mathbf{H}}_{NxM}^*$ is the complex conjugate transpose of $\tilde{\mathbf{H}}_{NxM}$. The channel matrix contains fading samples of the channels which means that we need to average \tilde{C} over the distribution of $\tilde{\mathbf{H}}_{NxM}$.

Normalization of Channel Matrix

The normalization of the channel matrix must be done with respect to the square root of the time average received power of a single reference antenna with 100% radiation efficiency located in the same environment, when the same total power is transmitted from the opposite side of the environment. The next subsection will show how typical channel matrices can be computed, and thereafter in Section 58.6 how they can be measured.

Numerical Simulation of Channels in Isotropic Fading Environment

Next is an explanation for simulating a fading environment numerically. First, distribute a set $\{E_k\}$ of $k = 1, 2, \dots, K$ plane wave sources with AoAs (θ_k, φ_k) randomly and uniformly distributed over a sphere surrounding the MIMO array. The sources E_k in the set have independent and normal-distributed complex amplitudes of the θ - and φ -polarized components. Calculate the received voltage V_i on each embedded antenna port i due to the set of such incident plane waves by using the equivalent circuit on reception:

$$V_i = -\frac{2j\lambda}{\eta I} \sum_{k=1,2,\dots}^K \mathbf{G}_i(\theta_k, \varphi_k) \cdot \mathbf{E}_k \quad (58-27)$$

where all parameters have been explained previously and $\mathbf{G}_i(\theta, \varphi)$ is the embedded far field function of port i . The power absorbed in the load R_L becomes

$$P_L = \frac{1}{2} R_L \left| \frac{V_i}{Z_a + Z_L} \right|^2 \quad (58-28)$$

which has its maximum value when $Z_L = Z_a^*$.

In order to determine the performance of a specific multiport antenna with N ports in an $M \times N$ MIMO system, it is assumed that the M ports at the opposite transmit side are completely uncorrelated. Next this is modeled by generating $M \times 1000$ sets of sources, with $k = 1, 2, \dots, 20$ sources in each set, and calculating $M \times 1000$ values of received voltages V_i .

The channel matrix \mathbf{H} should preferably be normalized in such a way that $|\mathbf{H}|^2$ represent power normalized to the average received power of a single ideal matched lossless reference antenna located in the same environment. In theoretical works we must be sure that this normalization is unaffected by possible differences in input impedances seen on the ports of the MIMO antenna and the reference antenna. This can be done by first introducing a normalized voltage

$$A_i = \sqrt{\frac{R_L}{2}} \frac{V_i}{(Z_a + Z_L)} \quad (58-30)$$

that by comparison with Eq. 58-28 is seen to represent the received power on port i when its absolute values is squared. Therefore, the appropriate normalization of the elements of the channel matrix \mathbf{H} becomes

$$H_i = A_i / \sqrt{P_{\max, \text{ref}}} \quad (58-31)$$

where $P_{\max, \text{ref}}$ is the maximum available average received power on the port of the reference antenna, i.e., the received power averaged over all $M \times 1000$ source configurations when we have corrected for the mismatch and ohmic losses in the reference antenna. In this way, the total radiation efficiency and mutual coupling of the elements of the MIMO antenna is included in the channel matrix \mathbf{H} . There are a total of $1000 \mathbf{H}_{M \times N}$ matrices in our above example of a $M \times N$ MIMO system in which values of V_i are found for each of the $M \times 1000$ source configurations and each of the N ports of the MIMO antenna under study.

$P_{\max, \text{ref}}$ can be found in the following way:

1. Expose a single reference antenna, for example, a single dipole, to the same sets of sources as the simulated radiation patterns of the embedded ports of the MIMO antenna.
2. Calculate the received voltage $V_{i, \text{ref}}$ at the port of the reference antenna using Eq. 58-27.
3. Determine the maximum available received power $p_{i, \max, \text{ref}}$ of the reference antenna using Eq. 58-28 with matched ports, i.e., $Z_L = Z_a^*$, where Z_a is the input impedance of the reference antenna, which gives

$$p_{i, \max, \text{ref}} = \frac{1}{2} R_a \left| \frac{V_{i, \text{ref}}}{2R_a} \right|^2 \quad (58-29)$$

4. Average this over all $M \times 1000$ sets of source configurations to get $P_{\max, \text{ref}}$.

Figure 58-9 shows an example of how the average maximum available spectral capacity varies with SNR for a circular array of monopoles with different spacing.

58.6 EMULATION OF MULTIPATH AND FADING USING REVERBERATION CHAMBER

The reverberation chamber is a large metal cavity provided with mode stirrers and one or more antennas, as shown in Figure 58-10. It has been used for more than 20 years for EMC measurements of electromagnetic susceptibility and radiated emissions of electronic devices.¹ The basic theories of it have also been well understood.¹⁴⁻¹⁷ The main EMC

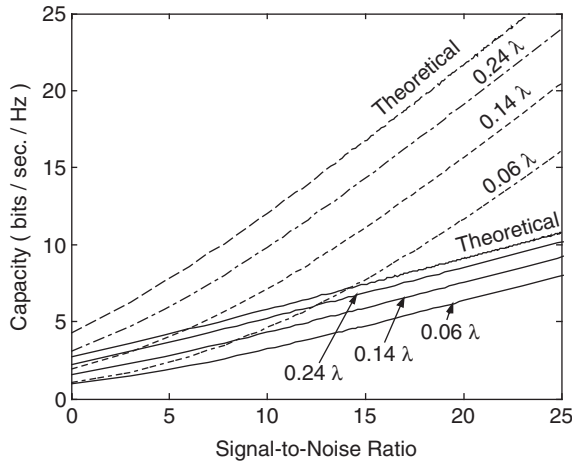


FIGURE 58-9 Example of average maximum available capacity for 6-port antenna in 1x6 (solid lines) and 3x6 MIMO systems (dashed lines). The multiport antenna consists of six monopoles located with equal distance in a circle. The distance between the neighboring monopoles were 0.24λ , 0.14λ , and 0.06λ . The “theoretical” curves represent ideal theoretical values for isolated independent MIMO elements.

application has been to generate high field strength for susceptibility testing. The last six years the reverberation chamber has been developed to a more accurate instrument for also measuring the characteristics of desired radiation of small antennas and active mobile terminals, in particular when these are intended for use in Rayleigh fading, such as for wireless/mobile communications in urban or indoor environments. Here we will describe how the reverberation chamber works, and how to measure the performance of antennas as defined in previous sections of this chapter.

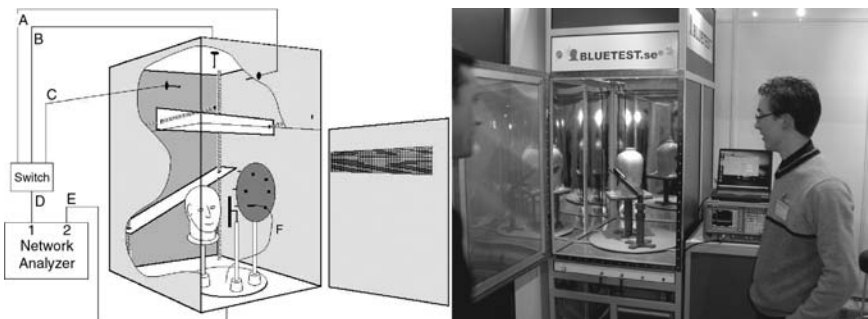


FIGURE 58-10 Drawing (left) and photo (right) of reverberation chamber, shown with open door.²⁰ The drawing shows the setup for measuring MIMO system capacity, and the photo shows the setup for measuring radiated power of mobile phone close to head phantom. The chamber size shown allows for use down to 850 MHz with 25 MHz frequency stirring.

Mode Stirring (Mechanical, Platform, and Polarization Stirring)

The basic measurement setup is illustrated in the left parts of both Figures 58-10 and 58-11. This is used both for calibration and measuring passive antenna performance. The transmission coefficient S_{21} is measured between two antennas in the chamber. One of these is the calibration antenna or the antenna under test (AUT), and the other is part of the chamber and could be wall-fixed. The chamber is so large that several modes are excited at the frequency of test, so the level of S_{21} may be larger than in free space, but it may also be lower, depending on how the modes are excited and how they combine (fading). Reverberation chambers are provided with different ways of stirring the modes and thereby the S_{21} level varies, i.e., emulation of fading. The most common mechanical stirring means are with fans or other large rotating or translating mechanical structures.

The chamber in Figure 58-10 has the following stirring capabilities:

- **Mechanical stirring** This is achieved by means of two plate-shaped mechanical stirrers, which can be moved along a complete wall and along the ceiling by means of step motors. The larger volume the stirrers cover, the better.
- **Platform stirring**¹⁸ The antenna under test is located on a rotatable platform which moves the antenna to different positions in the chamber. This stirring method is very effective in small chambers.
- **Polarization stirring**¹⁹ There are three orthogonal wall-fixed chamber antennas, and S_{21} is successively measured between each one of them and the AUT. Thereby, the chamber gets a good polarization balance.

There is also a fourth, very effective stirring method, referred to as *frequency stirring*. This corresponds to averaging S_{21} or $|S_{21}|^2$ over a frequency band during the processing of the results. This will be treated in the next section.

Measured S-parameters and Their Relation to Free Space Antenna Characteristics

Figure 58-11 illustrates how S_{11} and S_{21} are measured between two antennas inside a reverberation chamber. It can be physically argued that S_{11} of the AUT must consist

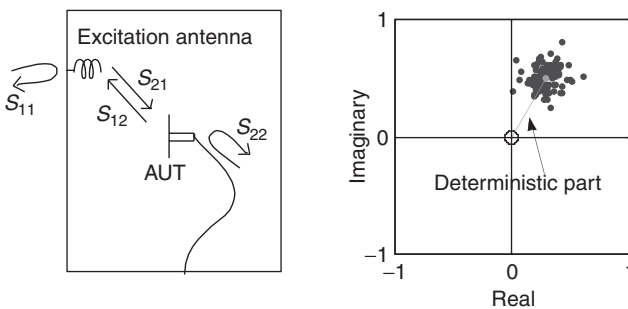


FIGURE 58-11 Example of measured S -parameters of two antennas in a reverberation chamber as a result of mode stirring, and an illustration of the deterministic and random contributions to them.

of two contributions; one contribution being the S_{11}^a from the antenna itself as if it was located in free space, and another S_{11}^c from the chamber:

$$S_{11} = S_{11}^a + S_{11}^c \quad (58-32)$$

The former of these contributions is deterministic, whereas the latter is random as a result of the stirring. If the number of independent values of S_{11}^c is large enough, it must get a complex Gaussian distribution with zero mean. Therefore, it is possible to determine S_{11}^a by complex averaging of the measured S_{11} over all stirrer positions, and if needed by additional complex frequency stirring:²¹

$$S_{11}^a = \frac{1}{M} \sum_{\substack{\text{stirrer} \\ \text{pos.}}} S_{11} = \overline{S_{11}} \quad (58-33)$$

where $\overline{S_{11}}$ is a compact way of writing the average of S_{11} . Similarly S_{22}^a of the wall-mounted chamber antenna can be found.

It can also be argued that S_{21} must consist of two contributions, one deterministic coupling S_{21}^d being the same as in free space, and another statistic contribution S_{21}^c coming from the chamber, i.e.,

$$S_{21} = S_{21}^d + S_{21}^c \quad (58-34)$$

Rayleigh Fading, Rice fading, and AoA Distribution

It is desirable that the reverberation chamber provide the isotropic reference environment with Rayleigh fading. Therefore, the direct coupling in Eq. 58-34 should be as low as possible. The direct coupling follows the free space transmission formula, also referred to as Friis transmission formula,⁴ which gives

$$|S_{21}^d|^2 = \left(\frac{\lambda}{4\pi r} \right)^2 G_t G_r \quad (58-35)$$

where r is the distance between the antennas, λ is the wavelength, and G_t and G_r are the realized gains of the wall-mounted antenna and AUT, respectively, in the direction of the opposite antenna. Therefore, the main lobes of the antennas should not point towards each other. The platform stirring is very effective to reduce direct coupling, as the direction of the AUT will be different in different platform positions. The direct coupling, together with loading the chamber, can also be used to get a controlled Rician distribution for test purposes.²²

From the discussion in the previous subsection, it is clear that the reverberation chamber provides Rayleigh fading if the direct coupling is reduced so much that it is much smaller than the chamber contribution to S_{21} . This can be done by platform stirring, or by using a wall-fixed antenna that has a null in the direction of the AUT. It was shown that each mode in a rectangular cavity can be expressed as eight plane waves, and that the directions of arrival of these plane waves are uniformly distributed over the unit sphere if there are enough modes excited.²³ Therefore, the reverberation chamber represents an isotropic multipath environment, i.e., the desired reference environment defined previously.

Average Transmission Level and Calibration

The transmission between two antennas in free space follows Friis transmission formula shown in Eq. 58-35. The corresponding formula for transmission between two antennas located in a reverberation chamber is Hill's transmission formula.¹⁶ This is valid when there is no direct coupling, and it presumes a large chamber with many excited modes. In the present text Hill's formula is referred to as the chamber transfer function and is presented in the following form:

$$G_{\text{chamber}} = \frac{P_r}{P_t} = \frac{c^3 e_{\text{rad}1} e_{\text{rad}2}}{16\pi^2 V f^2 \Delta f} \quad (58-36)$$

where f is the frequency, c is velocity of light, V is the chamber volume, $e_{\text{rad}1}$ and $e_{\text{rad}2}$ are the radiation efficiencies including mismatch of the two antennas, and Δf is the average mode bandwidth. The latter consists of four additive contributions due to wall losses, leakage, antennas in the chamber, and absorbing objects:

$$\Delta f = \sum_{\text{all walls}} \Delta f_{\text{wall}} + \sum_{\text{all slots}} \Delta f_{\text{leakage}} + \sum_{\text{all antennas}} \Delta f_{\text{antenna}} + \sum_{\text{all lossy objects}} \Delta f_{\text{object}} \quad (58-37)$$

with

$$\Delta f_{\text{wall}} = \frac{2A}{3V} \sqrt{\frac{c\rho f}{\pi\eta}}, \quad \Delta f_{\text{leakage}} = \frac{c\sigma_l}{4\pi V}, \quad \Delta f_{\text{antenna}} = \frac{c^3 e_{\text{rad}}}{16\pi^2 V f^2}, \quad \Delta f_{\text{object}} = \frac{c}{2\pi V} \sigma_a \quad (58-38)$$

and where V is the chamber volume, η the free space wave impedance, A the area of a conducting surface (such as a chamber wall) with surface resistance ρ , σ_l is the leakage cross section of a narrow slot in the chamber wall, and σ_a is the absorption cross section of an absorbing object. σ_l and σ_a ¹⁶ have slow frequency variation compared to the explicit frequency variation in the formulas. It should be noted that here the average chamber transfer function is expressed in terms of the average mode bandwidth rather than like Hill did in terms of the average Q of the modes.¹⁶ The reason is that the formula for Δf is much more compact because the different Δf contributions are additive, whereas Q contributions are not. Also, for specific chambers (at least loaded ones), the average mode bandwidth will not vary much with frequency, and therefore the value of Δf characterizes the chamber better than Q over a large frequency band. The total Δf s appearing in different practical chambers can be very different, corresponding to chamber Q s of between 30 and several thousands ($Q = f/\Delta f$). When measuring active terminals it is important that Δf is larger than the modulation bandwidth of the communication channel, otherwise measurement errors can appear.

In practice, a measurement of radiation efficiency, radiated power, or receiver sensitivity is based on first determining the chamber transfer function by calibration, using a reference antenna with known radiation efficiency $e_{\text{rad}2}$, with the AUT present in the chamber with its port matched terminated. The actual measurement will then be done by match-terminating the reference antenna and mounting the cable to the port of the AUT. The ratio between the two average transfer functions of the chamber in the two cases will be equal to the ratio between the radiation efficiency of the reference antenna and the AUT. The radiation efficiency $e_{\text{rad}1}$ of the chamber-fixed antenna (in this case wall mounted) does not need to be known, because it will be the same both when measuring the reference antenna and the AUT.

It is also possible to calibrate the chamber without having the AUT inside. However, the reference antenna must then also be removed when the AUT is measured. This is a simplified procedure, because the same calibration can be used when measuring several AUTs, and in particular it is simpler when measuring many active terminals. However, this simplified procedure only works if the chamber is loaded so much that the reference antenna and AUT does not represent any significant contribution to the total Δf of the chamber. If they do, the results of the measurements may be wrong, as the chamber transfer function no longer has a linear dependence on e_{rad} of the antenna. This fault may happen in particular if the reference antenna or AUT are made of materials which absorb radiation and thereby increase the total Δf even when the antenna ports are open or short-circuited.

Frequency Stirring on Mismatch-corrected Transfer Function

It has been found that if the free space mismatch efficiencies are removed from S_{21} in Eq. 58-36, the result is an average chamber transfer function that varies slower with frequency. This can then be frequency stirred for better accuracy without losing resolution due to variations in the mismatch efficiency. The free space input reflection coefficients of the two antennas can be obtained by complex averaging of the S_{11} and S_{22} measured in the reverberation chamber, as explained previously. Therefore, the mismatch-corrected chamber transfer function can be calculated using

$$G_{\text{chamber}} = \frac{1}{N} \sum_N \frac{|S_{21}|^2}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \quad (58-39)$$

where N is the number of stirrer positions. This function can be frequency stirred to improve accuracy.

Number of Independent Samples & Accuracy

In order to perform accurate measurements in the reverberation chamber, the chamber transfer function in Eq. 58-39 needs to be proportional to the radiation efficiency independent of which antenna or terminal is used, just like in Hill's theoretical transfer function in Eq. 58-36. This is possible only if the mode stirring creates enough independent samples. The S_{21} samples are complex normally distributed (for sufficient independent samples). Then, the relative accuracy by which we can estimate G_{chamber} has a standard deviation of⁵

$$\sigma = 1/\sqrt{N_{\text{ind}}} \quad (58-40)$$

where N_{ind} is the number of independent samples. This means that $N_{\text{ind}} = 100$ is needed for an accuracy of $\pm 10\%$, i.e., ± 0.5 dB. Thus, it is crucial that at least 100 independent samples are obtained by the mode stirring.

The number of independent samples is determined primarily by the mode density in the chamber—in other words, by the number of modes per MHz. This is a good approximation given by the classical formula

$$\partial N_{\text{mode}} / \partial f = V f^2 8\pi / c^3 \quad (58-41)$$

The number of independent samples is proportional to the mode density, but the proportionality constant is not known. It depends on chamber loading, mode stirring methods,

mechanical stirrer shapes, and chamber shape (at least for small chambers). The purpose of ongoing numerical studies is to determine these constants.²⁴⁻²⁵ A preliminary experimental study has been performed,²⁶ but it requires more work to get unique and simple conclusions. However, the following approximate relation between the mode density and number of independent samples can be preliminarily stated:

$$N_{\text{ind}} \leq 8[\partial N_{\text{mode}} / \partial f](\Delta f + B) \quad (58-42)$$

where B is the bandwidth of the frequency stirring and Δf is the same as in Eq. 58-36. The factor eight is due to platform stirring and is rooted both empirically and by physical reasoning. The latter argument goes as follows: Each cavity mode can be written as a sum of eight plane waves,²³ and the result is eight times more independent samples than modes. This is achieved by moving the antenna or terminal under test around in the environment, which is accomplished by platform stirring. The \leq sign in Eq. 58-42 means that this is an upper limit when there are enough samples and the mechanical mode stirring is sufficient.

It should be noted that the effect of the mechanical stirring can be so strong that N_{ind} will be larger than in Eq. 58-42. This will normally only appear when there is no or very small frequency stirring, and for high Q chambers, i.e., when Δf is small. Such conditions are common for EMC measurements, as high Q is needed in order to produce high field strength, so therefore, experiences with accuracy from EMC measurements are not directly transferable to accurate antenna and terminal measurements.

It is clear from Eq. 58-42 that the accuracy in a given chamber can be controlled by changing Δf and B . However, there may be other effects coming in when the loading is very large so some care must be taken. Still, a good Rayleigh distribution with many independent samples has been experienced even when the Q is as low as 30.²⁷ The problem with frequency stirring B is that the resolution becomes worse (resolution bandwidth increases). This means for antenna measurements that we cannot resolve variations in the radiation efficiency which are faster than B . In practice the resolution is somewhat better by using the mismatch correction to the transfer function before frequency stirring, like in Eq. 58-39. The reason is that mismatch efficiencies normally vary faster with frequency than efficiencies due to ohmic losses.

Normally, the chamber needs to be loaded in order to measure terminals of broadband systems, otherwise the modulation may be distorted. It has been shown that it is possible to measure systems with 25 MHz bandwidth in the chamber shown in Figure 58-10.

For smaller chambers the accuracy cannot be predicted based on Eqs. 58-41 and 58-42. The reason is that the mode density will not be accurate enough using Eq. 58-41 alone. The formula is approximate, and in reality the modes appear unevenly distributed as a function of frequency, depending on the shape of the chamber. However, the average mode density is quite well predicted by Eq. 58-41. The uneven mode distribution can be used to tune in a certain small chamber to good performance in a specific frequency band. This was done to determine the optimum height of a reverberation chamber for best accuracy in the GSM 900 transmit band.²³ To do this meant to actually count the number of modes for a rectangular box with the same size as the chamber that appears within a certain bandwidth B , instead of using Eq. 58-42. The frequency stirring values in Table 58-1 are based on such mode counting.

Accuracy and Resolution in Two Practical Chambers

The theoretical and empirical considerations in the previous sections have been used to generate the results for accuracy and resolution of two practical chambers with different sizes, as shown in Table 58-1. The results are obtained by counting modes based on the analytical formula for the resonance frequencies of modes in a rectangular cavity of the

TABLE 58-1 Performance of Two Different Reverberation Chambers (All values are empirical rooted in theoretical mode densities for a measurement accuracy of 0.5 dB STD.)

	Chamber 1	Chamber 2
Inner dimensions (DxHxW)	1.0 m × 1.75 m × 0.8 m	1.2 m × 1.75 m × 1.8 m
Isolation (shielding)	50 dB	100 dB
Lowest useful frequency	870 MHz with 25 MHz stirring	500 MHz with 20 MHz stirring
Resolution when meas. antennas*	circ. 12.5 MHz from 870 MHz	cir. 10 MHz from 500 MHz
Lowest frequency for full resolution	no frequency stirring needed above 2250 MHz	no frequency stirring needed above 1300 MHz
Resolution when measuring active terminals **		
GSM850/Cell TX 824-849	30 MHz, must average 6 channels 5 MHz apart, (not recommended)	10 MHz, must average 2 channels 5 MHz apart
GSM850/Cell RX 869 – 894	25 MHz, must average 5 channels 5 MHz apart	10 MHz, must average 2 channels 5 MHz apart
GSM900 TX 880 – 915	25 MHz, must average	10 MHz, must average
GSM900 RX 925 – 960	5 channels 5 MHz apart	2 channels 5 MHz apart
GSM1800 TX 1710 – 1785	10 MHz, must average	Single channel
GSM1800 RX 1805 – 1880	2 channels 5 MHz apart	
GSM1900 TX/PCS 1850 – 1910		
GSM 1900 RX /PCS 1930 – 1990		
Lowest frequency for full single channel resolution	2250 MHz	1300 MHz

* The resolution for antennas are a factor 2 better than the stirring bandwidth, due to access to S_{11} and S_{22} .
 **To get the specified 0.5 dB accuracy we must frequency stir by averaging results over channels as specified. The channels must be uniformly distributed over the specified resolution bandwidth. The chamber must be loaded to obtain an average mode bandwidth of about 4-5 MHz.

same size as the chamber. From this counting, the number of modes within different frequency stirring bandwidths is determined, as a function of frequency. Then, at a specific frequency of test, the stirring bandwidth needed to get more than 20 modes excited can be determined. Then, $N_{ind} > 8 \times 20 = 160 > 100$, which should be more than sufficient to get 0.5 dB RMS accuracy. In the case of measuring active phones, frequency stirring is obtained by averaging the results of several channels, as explained in Table 58-1.

58.7 MEASURING ANTENNAS AND ACTIVE TERMINALS IN REVERBERATION CHAMBER

Procedure for Calibration and Characterizing Multiport Antennas

The procedure for measuring a single- or multiport antenna in a reverberation chamber is briefly described as follows. The AUT is located inside the reverberation chamber in

such a way that it is more than 0.5 wavelengths from the walls and mechanical stirrers (for directive antenna longer distances will be needed in the main lobe direction). A single reference antenna with known radiation efficiency must be located far enough away from the AUT to avoid significant direct coupling (for non-directive antennas a spacing of half to one wavelengths is sufficient). One of the AUT ports is connected to a source, such as a network analyzer, and terminates all the other ports and the reference antenna in 50 Ohms. S -parameters are gathered between the port and the three wall-mounted antennas (used for polarization stirring) for all positions of the platform and the mechanical stirrers and for all frequency points. The measurement procedure is then repeated for every antenna port, also with the unconnected ports terminated in 50 Ohms, for exactly the same stirrer positions and position of the array inside the chamber. Thus, the field environment is exactly the same when measuring every port. The complex transmission coefficients S_{21} between the connected port and each of the three fixed wall antennas, as well as the reflection coefficients S_{11} of each of the wall antennas and S_{22} of the array port, are stored for every stirrer position and frequency point. Finally, the reference antenna is connected to the network analyzer and performs the same measurements as for the array. During the reference measurements, the AUT with all its ports terminated in 50 Ohms must be present in the chamber. This is necessary because the loading of the chamber (and thereby the Q -factor) needs to be the same during the measurements of both the reference antenna and the AUT, and because the AUT itself loads the reverberation chamber noticeably even when there is a lossy object such as a head phantom inside the chamber. In a small chamber it is advantageous to use frequency stirring (averaging) to improve accuracy. In such cases, the complex samples of S_{21} are corrected with mismatch factors due to both S_{11} and S_{22} before the frequency stirring, as shown in Eq. 58-39. The corrected S_{21} samples are normalized to the reference level corresponding to 100% radiation efficiency. This is obtained from the corrected S_{21} samples measured for the reference antenna, and its known radiation efficiency. These corrected and normalized samples of S_{21} are briefly referred to as the normalized S_{21} values. The normalized S_{21} values represent estimates of the channel matrix H of multipath communication channels set up between the wall antennas and the AUT inside the chamber. Therefore, from the measured S -parameters the diversity gain and capacity can be obtained as explained in previous sections.

In the previous explanations, the MIMO channel is set up between the chamber-fixed often wall-mounted antennas and the multiport AUT. It is of course also possible to locate two multiport antennas inside the chamber and set up the MIMO channels between the ports of these two antennas, and evaluate them as a complete MIMO system.

Radiated Power, Active Implemented Diversity Gain, and Receiver Sensitivity

Complete mobile terminals like cell phones can also be measured in reverberation chambers. The total radiated power is readily measured by first finding the chamber transfer function in Eq. 58-39 by using a reference antenna with known efficiency, and thereafter measuring the radiated power from the phone by connecting a spectrum analyzer to the chamber-fixed antennas.²⁸⁻²⁹ The radiation from the phone must be controlled by using a base station emulator or mobile communication test instrument. Phones with implemented diversity antennas can also be measured to determine the actual implemented diversity gain.³⁰

Even more important for wireless internet and multimedia terminals is the receiver sensitivity, as it will directly affect the time for downloading data. The receiver sensitivity is characterized by the signal level needed to reach a certain specified bit error rate (BER) or frame error rate (FER), the latter for CDMA systems. The limiting static signal level is, when averaged over many directions in the radiation pattern, referred to as the total isotropic sensitivity (TIS). An alternative in reverberation chambers is to measure BER/FER during

continuous fading to determine what is referred to as average fading sensitivity (AFS).³¹ The AFS can also be used to determine the TIS, as there is a system specific relation between the two values. The AFS is a more realistic performance parameter than TIS, and, it is much faster to measure.

ACKNOWLEDGMENTS

Many thanks to Andreas Wolfgang, Jan Carlsson, and Daniel Nyberg for their corrections and comments to the manuscript.

REFERENCES

1. M. Bäckström, O. Lundén, and P.-S. Kildal, "Reverberation Chambers for EMC Susceptibility and Emission Analyses," *Review of Radio Science 1999-2002* (2003): 429–452.
2. William H. Tranter (ed.), Brian D. Woerner (ed.), Theodore S. Rappaport (ed.), Jeffrey H. Reed (ed.), *Wireless Personal Communications: Channel Modeling and Systems Engineering* (Norwell, MA: Kluwer Academic Publisher, 2000).
3. R. Vaughan and J. B. Andersen, *Channels, Propagation and antennas for Mobile Communications* (London: The Institution of Electrical Engineers, 2003).
4. P.-S. Kildal, *Foundations of Antennas—A Unified Approach* (Lund, Sweden: Studentlitteratur, 2000).
5. T. Taga, "Analysis for Mean Effective Gain of Mobile Antennas in Land Mobile Radio Environments," *IEEE Trans. Vehicular Technology*, no. 39 (May 1990): 117–131.
6. P.-S. Kildal and K. Rosengren, "Correlation and Capacity of MIMO Systems and Mutual Coupling, Radiation Efficiency and Diversity Gain of their Antennas: Simulations and Measurements in Reverberation Chamber," *IEEE Communications Magazine*, vol. 42, no. 12 (December 2004): 102–112.
7. R. C. Hansen, *Microwave Scanning Antennas: Vol II, Array Theory and Practice* (New York: Academic Press 1964).
8. K. Rosengren and P.-S. Kildal, "Radiation Efficiency, Correlation, Diversity Gain, and Capacity of a Six Monopole Antenna Array for a MIMO System: Theory, Simulation and Measurement in Reverberation Chamber," *Proceedings IEE, Microwave Antennas Propagation*, vol. 152, no. 1 (February 2005): 7–16. (See also Erratum published in August 2006.)
9. P.-S. Kildal and K. Rosengren, "Electromagnetic Analysis of Effective and Apparent Diversity Gain of Two Parallel Dipoles," *IEEE Antennas and Wireless Propagation Letters*, vol. 2, no. 1 (Jan., 2003): 9–13.
10. C. Ludwig, "Mutual Coupling, Gain and Directivity of an Array of Two Identical Antennas," *IEEE Trans. Antennas Propagat.* (November 1976): 837–841.
11. P. W. Hannan, "The Element-Gain Paradox for a Phased-Array Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-12, no. 7 (July 1964): 423–433.
12. P.-S. Kildal, K. Rosengren, J. Byun, and J. Lee, "Definition of Effective Diversity Gain and How to Measure It in a Reverberation Chamber," *Microwave and Optical Technology Letters*, vol. 34, no. 1, (July 2002): 56–59. R. G. Vaughan, and J. B. Andersen, "Antenna Diversity in Mobile Communications," *IEEE Trans. Veh. Technol.*, vol. VT-36 (November 1987) 149–172.
13. D. A. Hill, "Electronic Mode Stirring for Reverberation Chambers," *IEEE Transactions on Electromagnetic Compatibility*, vol. 36, no. 4 (November 1994): 294–299.
14. M. Schwartz, W. R. Bennett, and S. Stein, *Communication System and Techniques* (New York: McGraw-Hill, 1965).
15. J. G. Kostas and B. Boverie, "Statistical Model for a Mode-Stirred Chamber," *IEEE Transactions on Electromagnetic Compatibility*, vol. 33, no. 4 (November 1991): 366–370.

16. D. A. Hill, M. T. Ma, A. R. Ondrejka, B. F. Riddle, M. L. Crawford, and R. T. Johnk, "Aperture Excitation of Electrically Large, Lossy Cavities," *IEEE Transactions on Electromagnetic Compatibility*, vol. 36, no. 3 (August 1994): 169–178.
17. D. A. Hill, "Linear Dipole Response in a Reverberation Chamber," *IEEE Transactions on Electromagnetic Compatibility*, vol. 41, no. 4 (November 1999): 365–368.
18. K. Rosengren, P.-S. Kildal, C. Carlsson, J. Carlsson, "Characterization of Antennas for Mobile and Wireless Terminals in Reverberation Chambers: Improved Accuracy by Platform Stirring," *Microwave and Optical Technology Letters*, vol. 30, no. 20 (September 2001): 391–397.
19. P.-S. Kildal and C. Carlsson, "Detection of a Polarization Imbalance in Reverberation Chambers and How to Remove it By Polarization Stirring when Measuring Antenna Efficiencies," *Microwave and Optical Technology Letters*, vol. 34, no. 2 (July 2002): 145–149.
20. P.-S. Kildal, "A Method and an Apparatus for Measuring the Performance of Antennas, Mobile Phones, and Other Wireless Terminals," International Patent Application (31 March 2000): CT/SE01/00422.
21. P.-S. Kildal, C. Carlsson, and J. Yang, "Measurement of Free Space Impedances of Small Antennas in Reverberation Chambers," *Microwave and Optical Technology Letters*, vol. 32, no. 2 (January 2002): 112–115.
22. C. L. Holloway, D. A. Hill, J. M. Ladbury, P. Wilson, G. Koepke, and J. Coder, "On the Use of Reverberation Chambers to Simulate a Controllable Rician Radio Environment for the Testing of Wireless Devices," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, (November 2006): 3167–3177.
23. K. Rosengren and P.-S. Kildal, "Study of Distributions of Modes and Plane Waves in Reverberation Chambers for Characterization of Antennas in Multipath Environment," *Microwave and Optical Technology Letters*, vol. 30, no. 20 (September 2001): 386–391.
24. U. Carlberg, P.-S. Kildal, and J. Carlsson, "Study of Antennas in Reverberation Chamber Using Method of Moments with Cavity Green's Function Calculated By Ewald Summation," *IEEE Trans. Electromagn. Compat.*, vol. 47, no. 4 (November 2005): 805–814.
25. K. Carlsson, J. Carlsson, and P.-S. Kildal, "A Method of Moments Solution for Wires Inside a Reverberation Chamber Using Spectral Domain Techniques and Asymptote Extraction," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, (November 2006): 3106–3113.
26. Marzari, Eric, "Physical and Statistical Models for Estimating the Number of Independent Samples in the Chalmers Reverberation Chamber," Master's thesis, Chalmers, 2004.
27. C. Orlenius, M. Franzen, P.-S. Kildal, and U. Carlberg, "Investigation of Heavily loaded Reverberation Chamber for Testing of Wideband Wireless Units," *IEEE AP-S International Symposium*, Albuquerque, NM, July 2006.
28. P.-S. Kildal and C. Carlsson (C. Carlsson now has the name C. Orlenius), "TCP of 20 Mobile Phones Measured in Reverberation Chamber: Procedure, Results, Uncertainty and Validation," Bluetest AB (February 2002): www.bluetest.se.
29. N. Serafimov, P.-S. Kildal, and T. Bolin, "Comparison Between radiation Efficiencies of Phone Antennas and Radiated Power of Mobile Phones Measured in Anechoic Chambers and Reverberation Chamber," *IEEE AP-S International Symposium*, San Antonio, Texas, June 2002.
30. R. Bourhis, C. Orlenius, G. Nilsson, S. Jinstrand, and P.-S. Kildal, "Measurements of Realized Diversity Gain of Active DECT Phones and Base-Station in a Reverberation Chamber," *IEEE AP-S International Symposium*, Monterey, CA, June 2004.
31. C. Orlenius, P.-S. Kildal, and G. Poilasne, "Measurements of Total Isotropic Sensitivity and Average Fading Sensitivity of CDMA Phones in Reverberation Chamber," *IEEE AP-S International Symposium*, Washington, D.C., July 3–8, 2005.

Chapter 59

Computational
Electromagnetics
for Antennas

Leo C. Kempel

Michigan State University

CONTENTS

59.1	INTRODUCTION.....	59-2
59.2	TIME-DOMAIN METHODS.....	59-5
59.3	FREQUENCY-DOMAIN METHODS.....	59-9
59.4	SUMMARY OF PUBLIC DOMAIN AND COMMERCIAL CODES: <i>CAVEAT EMPTOR</i>	59-20
59.5	SOURCES FOR PARALLEL PROGRAMMING INFORMATION.....	59-23

59.1 INTRODUCTION

Computational electromagnetics (CEM) has perhaps had the greatest impact on antenna design than any technology since the invention of the cavity magnetron. That is a strong statement; however, it is now common for antenna designers to make use of either commercial or custom-design CEM codes in nearly all design cycles. Indeed, calculated performance is often required in addition to measured results! These tools are used to predict performance as well as to better understand the sometimes complex physical interactions that occur between the antenna itself, the source of radio frequency (RF) signal, and the structure in the vicinity of the antenna. Traditional antenna design considered the antenna in isolation (for example, an idealized feed and free space environment) and made assumptions necessary for determination of engineering significant quantities such as the voltage standing wave ratio (VSWR), driving-point impedance, efficiency, axial ratio, and directivity, amongst others. CEM methods, both computationally exact ones based on a full-wave solution of Maxwell's equations and approximate ones based on some physically justifiable assumptions, have considerably improved design methodologies available to antenna engineers. Namely, these CEM tools allow much greater fidelity for the designer so that the goal of a virtual design environment with sufficient accuracy and included physics is within reach during the first decade of the 21st century.

Examples of improved fidelity include the addition of surrounding structure, realistic feed models, non-trivial materials (for example, dispersive, dissipative, anisotropic, and so on), finite conductive metals, and even active and/or passive electronic components. It stands to reason that the environment in the vicinity (within the near-field) of an antenna will have a strong impact on the performance of that device. Nevertheless and by necessity, antennas were traditionally designed assuming free space. This considerably simplified the mathematics; however, often designers found dramatically altered performance, especially for near-field quantities such as VSWR, when the antenna was placed in its operating environment (for example, on an automobile, aircraft, ship, or in the vicinity of buildings). Likewise, the interactions between the feed and electronics within the antenna structure itself can dramatically impact actual versus predicted performance. Since one of the main rationales for investing in a CEM design tool is the reduction—and even elimination—of costly prototype testing, improved fidelity is obviously necessary for instilling confidence in the computed results. Finally, traditional antenna designs have either approximated or ignored the impact of dielectric or magneto-dielectric materials on antenna performance. With the introduction of advanced CEM design tools these assumptions are no longer necessary.

The tremendous advances in computer resources over the past several decades have made the inclusion of these important physical mechanisms possible. An impact of Moore's Law, "doubling the number of transistors on an integrated circuit every eighteen months," is that computational power increased exponentially. CEM, among many other computational science and engineering disciplines, has reaped the benefits of the power law growth in affordable computing power, primarily via the inclusion of greater fidelity. With the emergence of commercially available multicore processors in 2005, the power of parallel programming, as applied to CEM methods, will become commonly available allowing the inclusion of even more physical mechanisms.

The scientific literature over the past 20 years or so is rich in advances made in computational methods that can be (and often are) applied to antenna design and analysis. It is not the intent of this chapter to give a detailed list of the most important papers; doing so would rapidly consume the chapter resulting in a bibliography rather than a practically useful introduction to CEM methods for antenna design and analysis. Rather, the cited references will primarily be a subset of the books and journal papers that have emerged over the past decade

or so usually based on scientific papers, conference publications, and technical reports. The interested reader can go to the references in those texts for primary sources.

Abbreviated Summary of Progress since the 1960s

A brief summary of the progress in computational modeling for antennas is in order prior to proceeding. There are, arguably, five major advances in CEM that have had a profound impact on antenna design. These are (in roughly chronological order):

- The development of the finite difference-time domain (FD-TD) method
- The release of the Numerical Electromagnetics Code (NEC) computer program by Lawrence Livermore National Laboratory (LLNL)
- The discovery of the Rao-Wilton-Glisson (RWG) basis function
- The development of three-dimensional vector finite elements
- The discovery of fast solver acceleration methods

A brief summary of each of these items is provided to give the reader a feeling for the interplay between research discovery and its impact on antenna design.

The FD-TD method really became a practical CEM tool with the development of the Yee cell in 1966.¹ As will be discussed in detail in the next section, the point-form of Maxwell's equations are a set of continuous, coupled, partial differential equations in both space and time. To solve these equations on a digital computer, a method for approximating these exact equations with differences in both space and time is needed. The Yee cell provided a construct for doing so, in that both the electric field and magnetic field are represented spatially with offset grids, while temporally, they are represented half a time step apart. This allowed for the specification of second-order recurrence relationships thereby reducing the continuous partial differential equation to a manageable set of difference equations implemented as recurrence expressions. The result is a practical method for modeling three-dimensional structures in the time-domain; one of the most efficient methods for wide bandwidth antenna study.

The release of the frequency-domain integral equation program NEC-1 in 1980² initiated a flurry of activity in utilization of a CEM design tool for antennas. Previous to NEC, various research organizations (primarily universities but also research laboratories with a variety of affiliations) developed and utilized CEM codes for antenna design and analysis. However, these codes were proprietary and thus had very limited market penetration. Since LLNL made NEC widely available (indeed NEC-2 is freely available worldwide and has spawned a variety of support program packages, some of which are listed later in this chapter), it dramatically impacted antenna design. NEC utilizes wire and surface patches in defining the geometry and over which the unknown currents are computed.

Surface patches were one of the more challenging aspects of modeling antennas with an integral equation formulation. The reason for this is that the expansion functions at the time did not correctly model current continuity from patch-to-patch (as compared to the wire elements in NEC that did indeed enforce current continuity at wire junction nodes). The development of vector basis functions by Rao, Wilton, and Glisson (RWG) in 1982³ resulted in a flurry of new integral equation codes, both commercial and non-commercial. The inclusion of surface currents allowed simulation of an entirely new class of antennas and improved modeling of the surface structure in the vicinity of the antenna. Heretofore, the environment in the vicinity of an antenna was modeled with a collection of wire grid elements rather than a solid surface. The results of using such a wire grid model were

sometimes acceptable and sometimes not. This meant that the introduction of the RWG basis function was a substantial improvement.

Most integral, equation-based design programs were suitable for simulation of perfect electric conductors, and with some approximation, imperfect conductors in free space. They were not able to efficiently model dielectric antenna loading materials. FD-TD design codes could model materials, but being time-domain formulations, FD-TD codes are not efficient for simulation of resonant antennas. The finite element (FE) method, long a staple in the computational fluid dynamics and the computational heat transfer communities, had the potential for including materials. However there were two serious limitations that had to be overcome: (1) the need for an electromagnetically correct vector expansion function and (2) the need to efficiently truncate the computational domain for open problems as is the case for an antenna. The first limitation was overcome with the development of a suitable basis function,⁴ while the second was solved for a class of problems by combining the finite element method with a boundary integral (the so-called FE-BI method).⁵ Later, more accurate local mesh truncation conditions became important for commercial applications, due to the generally lower memory requirements as compared to a boundary integral termination. These two innovations meant that for the first time, the interactions between a material and the antenna can be rigorously included in the model, yet do not require an excessive computational burden. (Note: In 1990, the computational cost of a volume integral equation method was too high to be practical, but in the 21st century that burden can be accommodated for practical simulation tasks with high performance computers.)

The final trend-setting innovation for antenna modeling was the introduction of *fast solvers*. These are based on one of a handful of observations. In one incarnation—the fast multipole-based methods (FMM)—the interactions between elements are computed from the interactions between local groups. This reduces the computational burden, both storage and computation, depending on how many levels are included in the multipole method⁶ for both frequency-domain and time-domain formulations. In another popular implementation of a fast method, based on different assumptions and mathematics, the interactions between the unknowns are computed based on the results, using auxiliary unknowns. The computation of these values is much more efficient than a brute-force approach, using RWG functions due to the uniform spatial arrangement of these auxiliary unknowns. For some important antenna structures, the use of this Adaptive Integral Method (AIM) can dramatically increase the efficiency of a simulation through the use of fast Fourier transforms (FFT) during the calculation of matrix-vector products within an iterative matrix solver.⁷ The impact of these efficient solution techniques is a dramatic increase in the electrical size of a geometry that can be practically computed within a design time-frame.

The remainder of the chapter is divided in four parts:

- Time-domain methods
- Frequency-domain methods
- A brief survey of commercial and public domain computer programs
- A brief discussion of parallel programming methods and sources

The first two categories represent the bulk of the material in this chapter, and they are further subdivided into solution techniques: finite difference, integral equation, finite element, and alternative methods as appropriate. The reader will notice that much more space in the chapter is devoted to a discussion of the frequency-domain methods as compared to time-domain methods. There are two reasons for this: (1) as of publication of this edition of the handbook, frequency-domain CEM tools are more widely used than time-domain methods for a variety of reasons (not the least of which is that most antennas are resonant structures rather than wide bandwidth radiators) and (2) the author is more familiar with

frequency-domain methods. With the increased importance and market penetration of wide bandwidth antennas, time-domain simulators will become increasingly more important to the community.

59.2 TIME-DOMAIN METHODS

Formulation of Maxwell's equations in the time-domain is the most natural approach, for the simple reason that all physical electric and magnetic fields are time-dependent. As stated previously, Maxwell's equations are a set of coupled, first-order, time-dependent expressions that originally were presented in integral form

$$\begin{aligned}
 \oint_C \mathbf{E} \cdot d\mathbf{L} &= -\frac{\partial}{\partial t} \int_S \mu \mathbf{H} \cdot d\mathbf{S} \\
 \oint_C \mathbf{H} \cdot d\mathbf{L} &= \frac{\partial}{\partial t} \int_S \epsilon \mathbf{E} \cdot d\mathbf{S} + \int_S \mathbf{J} \cdot d\mathbf{S} \\
 \oint_{\partial V} \epsilon \mathbf{E} \cdot d\mathbf{S} &= \int_V \rho_e dV \\
 \oint_{\partial V} \mu \mathbf{H} \cdot d\mathbf{S} &= 0
 \end{aligned} \tag{59-1}$$

In this, the electric field (\mathbf{E}), magnetic field (\mathbf{H}), and current density (\mathbf{J}) are three-dimensional vector fields that are dependent on both time and space. The material properties are denoted by the permittivity (ϵ) and permeability (μ). For simplicity, an isotropic material is assumed for these expressions; however, in the case of an anisotropic material (such as ferrimagnetic material, namely saturated yttrium iron garnet or YIG), a suitable formulation can be derived that includes the anisotropy of the material. The integration domains are over a spatial volume (V), the boundary surrounding that volume (∂V), an open surface (S), and the contour that bounds that open surface ($C = \partial S$). Finally, the electric charge density (ρ_e) is, like the current density, a source of electromagnetic fields and is a function of both time and space. Although there are solutions that utilize the integral form of Maxwell's equation, most notably the Finite Integral Technique (FIT),⁸ most time-domain Maxwell equation solvers utilize the point-form of the equations. These are given by

$$\begin{aligned}
 \nabla \times \mathbf{E} &= -\frac{\partial}{\partial t} (\mu \mathbf{H}) \\
 \nabla \times \mathbf{H} &= \frac{\partial}{\partial t} (\epsilon \mathbf{E}) + \mathbf{J} \\
 \nabla \cdot (\epsilon \mathbf{E}) &= \rho_e \\
 \nabla \cdot (\mu \mathbf{H}) &= 0
 \end{aligned} \tag{59-2}$$

These are a set of first-order, coupled partial differential equations (PDE) with two vector-valued unknown fields: \mathbf{E} and \mathbf{H} . There are three prevailing methods for solving these equations besides the aforementioned FIT approach: finite differences, integral equations, and finite element methods. These will each be discussed with minimal detail.

Finite Difference-Time Domain (FD-TD) Method

The FD-TD method is the most widely developed and popular time-domain Maxwell equations solver. There is a plethora of codes available (both commercial and in the public domain) that can be utilized for antenna design and analysis. See the end of

Section 59.4 for a brief discussion of these sources. Many people believe that the FD-TD method is the most natural way to solve Maxwell's equation. With this method differential operators are approximated using difference operators resulting in both the electric and magnetic field as a function of time, for example, a direct solution of Eq. 59-2. However, the reader should be cautioned that the most prevalent implementations of the FD-TD method (those using leap-frog methods) do not provide both the electric and magnetic fields at the same instant in time. Also, these fields are not provided at the same point in space, since the solution is given over two offset grids such as with the Yee-cell case.¹ Nevertheless, FD-TD solutions are popular for at least two reasons: (1) They provide considerable physical insight into the radiation mechanism of an antenna when the method is combined with a visualization capability and (2) it is relatively easy to program a simple two-dimensional and even three-dimensional FD-TD program. (This is not to say that it is easy to write efficient code or to provide many of the necessary features—such as dispersive media and accurate feed models—that are now incorporated in commercial codes.)

The simplest FD-TD formulation to consider is the one-dimensional case. In this case, all waves are plane waves that travel along the single dimension, parallel to the x -axis. For the one-dimensional case, Maxwell's point form equations reduce to⁹

$$\begin{aligned} \frac{\partial H_y}{\partial t} &= \frac{1}{\mu} \frac{\partial E_z}{\partial x} \\ \frac{\partial E_z}{\partial t} &= -\frac{1}{\varepsilon} \frac{\partial H_y}{\partial x} \end{aligned} \quad \text{TM polarization} \quad (59-3)$$

or

$$\begin{aligned} \frac{\partial E_y}{\partial t} &= -\frac{1}{\varepsilon} \frac{\partial H_z}{\partial x} \\ \frac{\partial H_z}{\partial t} &= -\frac{1}{\mu} \frac{\partial E_y}{\partial x} \end{aligned} \quad \text{TM polarization} \quad (59-4)$$

Note that in Eqs. 59-3 and 59-4, the material constitutive parameters, ε and μ , are not relative terms; it is a common mistake to use material parameters that are relative to free space. In addition, since these are time-dependent field quantities, the constitutive parameters are real-valued. Loss is represented by a conductivity that has been suppressed in Eqs. 59-3 and 59-4 for clarity's sake. To solve these coupled, first-order equations, the differences will be replaced with second-order central differences. For example, the PDE $\partial H/\partial t = (\mu^{-1})\partial E/\partial x$ is approximated as

$$\left. \frac{H\left(t_n + \frac{\Delta t}{2}\right) - H\left(t_n - \frac{\Delta t}{2}\right)}{\Delta t} \right|_{x=x_i} = \frac{1}{\mu_i} \left. \frac{E\left(x_i + \frac{\Delta x}{2}\right) - E\left(x_i - \frac{\Delta x}{2}\right)}{\Delta x} \right|_{t=t_n} \quad (59-5)$$

where the time step is indicated by (Δt) and the spatial increment is given by Δx . Note that the permeability is evaluated at the i th spatial sample location, x_i . The n th time step is denoted by t_n . This difference equation (Eq. 59-5) is implemented as a recurrence relation as

$$H\left(t_n + \frac{\Delta t}{2}\right) \Big|_{x_i} = H\left(t_n - \frac{\Delta t}{2}\right) \Big|_{x_i} + \frac{\Delta t}{\mu \Delta x} \left[E\left(x_i + \frac{\Delta x}{2}\right) - E\left(x_i - \frac{\Delta x}{2}\right) \right] \Big|_{t_n} \quad (59-6)$$

or more typically written as

$$H_i^{n+1/2} = H_i^{n-1/2} + \frac{\Delta t}{\mu \Delta x} [E_{i+1/2}^n - E_{i-1/2}^n] \quad (59-7)$$

where the superscript indicates the instance in time and the subscript indicates the spatial location of the grid point. Notice that the electric field is not ever evaluated at the same point in time or space as the magnetic field, and vice versa. The companion formula for the electric field is given by

$$E_{i+1/2}^{n+1} = E_{i+1/2}^n + \frac{\Delta t}{\epsilon \Delta x} [H_{i+1}^{n+1/2} - H_i^{n+1/2}] \quad (59-8)$$

The fields are evaluated as time increases by alternating between these two equations, increasing time a half step between use of either Eq. 59-7 or Eq. 59-8. This is graphically depicted in Figure 59-1.

One of the issues that must be considered in any time-domain solution method is its stability. Specifically, whether the algorithm is guaranteed not to generate non-physical, increasing fields as time progresses. The core issue is the relationship between the time step (Δt) and the spatial step (Δx). If the time step is too long, then a field can pass through more than one cell during a time step as it propagates, generating an unstable solution. For the one-dimensional case, the so-called magic time step¹⁰

$$\Delta t = \frac{\Delta x}{c} \quad (59-9)$$

can be defined as the results in a grid velocity (the velocity of propagation within the numerical grid used to approximate real space) that *exactly* reproduces the velocity of propagation in real space with c being the speed-of-light in a vacuum. Because of this, grid dispersion (the phenomena that a signal propagates in the discretized space at a different velocity than in real space) does not occur. In Eq. 59-9, if the magic time step is

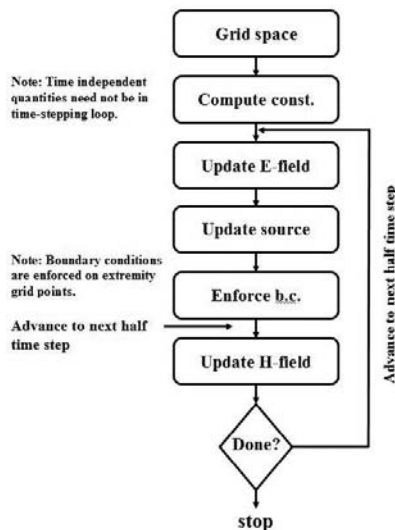


FIGURE 59-1 Flow-chart for FD-TD algorithm

not used, the time step must be less than the spatial step divided by the highest speed of propagation to maintain stability. In all other dimensions, the grid velocity only approximately represents the true continuous domain velocity and therefore numerical dispersion will occur. The Courant Stability Criterion for two- and three-dimensional grids is given in many texts.⁹⁻¹¹

As an example of the power inherent in FD-TD simulations, consider the temporal snapshot of a plane wave incident on a multilayered dielectric shown in Figure 59-2. In this two-dimensional simulation, the Gaussian plane wave is injected at the left boundary of the figure and propagates to the right. The simulation was halted as the wave was partially reflected by the first material discontinuity (located at $x = 0.15$) and partially transmitted at that boundary. As seen, one of the powerful features of any time-domain simulation is the ability to stop the simulation to capture relevant physics. Another powerful feature of time-domain simulations is the physical insight gained when the FD-TD simulation includes graphics. Notice that in this case, the reflected field is inverted while the transmitted field is not (as required by the transition conditions at the material discontinuity). Examples of using FD-TD for antenna analysis—both far-zone quantities such as directivity and near-zone quantities such as VSWR—are given.⁹ The MATLAB-code used to create Figure 59-2 is provided on the Internet (directions for obtaining all the computer programs used in here can be found at the end of the chapter). The computational cost of an FD-TD simulation can be large, especially if the grid is dense (a very large number of spatial points) and the signal is resonant. For a resonant signal, the ring-down time is very long and so the simulation must be run for many time steps. FD-TD, fortunately, is well-suited for parallelization, allowing the power of multiple processors to be brought to bear on a computationally intense problem. That is not to say that parallelization is trivial—domain decomposition and sharing of domain boundary information must be done well. However, it is somewhat more amenable to parallelization as compared to other methods (such as the FMM), which require very specialized parallelization skills.

Time Domain-Integral Equation (TD-IE) Method

The time domain-integral equation method solves integral equations directly in the time-domain. It contains the same essential elements of the FD-TD method (spatial discretization using basis function and some type of time stepping procedure). TD-IE was a topic of relatively early investigation (see the example of transient radiation by wire antennas¹²);

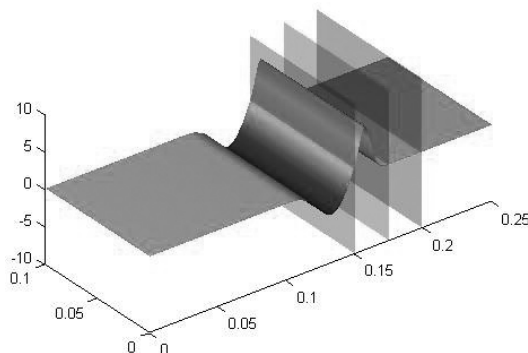


FIGURE 59-2 Snapshot of a plane wave striking a multilayer surface (Compliments of Mr. Steven Cossmann)

however, it suffered from stability issues that quite frankly are an issue of continuing concern. Recently, the CEM community has made substantial progress in both the flexibility of TD-IE methods and their stability. For example, surfaces are now represented using surface expansion functions rather than a porous collection of wires. The key to improving stability lies in the use of advanced expansion methods and representations in addition to very accurate calculation of integral quantities. Examples of recent innovations are given in the reference section.¹³⁻¹⁶ Note that most of the published work indicates that TD-IE methods were initially developed for scattering problems; however, they are being adapted for radiation and electromagnetic compatibility (EMC) applications.

Time Domain-Finite Element (TD-FE) Method

Another powerful time-domain CEM method is one that solves the partial differential equations (such as Eq. 59-2) using a finite element procedure rather than finite differences. Typically, these methods are referred to as the finite element-time domain (FE-TD) method. They are loosely related to the finite volume-time domain (FV-TD) method.¹⁷ The TD-FE method is undergoing continued development and is, in some ways, the most immature of CEM techniques described in this chapter. In 1997, a review paper¹⁸ discussed the state-of-the-art which has undergone considerable improvement since that time.¹⁹⁻²¹ One paper that is particularly interesting describes a hybrid method combining the best features of FD-TD, FE-TD, and TD-IE.¹⁹ The field quantities are functions of both time and space. Hence, the spatial field variation is represented by basis functions that have support only over limited portions of the geometry (typically only a few, adjacent finite elements). These are so-called subdomain basis functions. The temporal response is typically represented using temporal expansion functions. Both explicit and implicit methods have been proposed in the literature. Based on the power of the method—time-domain simulations with very flexible geometrical and material capabilities—and recent activity in the scientific literature, it can be expected that the TD-FE method will become an important, and possibly major, component in any complete CEM tool suite for antenna applications.

Additional Time-Domain Methods

Other widely used time-domain methods include implementation of the finite integration technique (FIT).²² This method, as briefly mentioned previously, implements solutions of the integral form of Maxwell's equations. FIT is widely used in designing the linear accelerator cavities; however, it is also finding increased use in antenna design. Undoubtedly there are more time-domain methods than those presented previously (for example, time-domain versions of the Uniform Theory of Diffraction²³). With apologies to the many contributions made around the world, this section focused on the most widely used and developed approaches.

59.3 FREQUENCY-DOMAIN METHODS

Two of the most frequently used frequency-domain methods are the Method of Moments and the Finite Element Method. In this section, each of these methods is presented in some detail although the interested reader should consult one of the many excellent textbooks on the subject. Other frequency-domain methods are introduced in a cursory manner at the end of this section.

Method of Moments (MoM)

Integral equations formulated and solved in the frequency-domain, colloquially referred to as the Method of Moments (MoM), is one of the most popular and widely used methods for antenna design and analysis. From the first major textbook on the subject,²⁴ the MoM method has been improved in terms of capability and fidelity since the 1960s. Other terms commonly used for these solution methods are integral equations and boundary element methods, among others.

A typical feature in the derivation and solution of such formulations involves the specification of an appropriate Green's function in the formulation of integral relations. These integral relations are then used to enforce appropriate boundary conditions resulting in an integral equation. The conversion of a continuous domain integral equation into a linear system of equations involves a discretization process, commonly known as the method of moments. One consequence of the use of a Green's function, with the resulting global communication between elements (also known as unknowns), is the resulting fully populated matrix. In general, the storage requirement of a MoM formulation scales as $O(N^2)$ where N is the number of degrees-of-freedom (unknowns). This can severely limit the scaling potential of a MoM-based computer program with increasing operating frequency or model complexity. For example, the storage requirement for a fully-populated matrix developed using a full three-dimensional current formulation, such as you might have for a volume integral equation, will scale to the sixth power! So, for a given physical problem, a doubling of the operating frequency will require eight times the number of unknowns and the matrix storage will be 64 times greater.

There is a plethora of excellent sources of information concerning integral, equation-based methods for simulating the performance of antennas. There are two recent books^{25–26} that are particularly useful for novice CEM users, because they were written to complement the commonly available MoM computer programs. Note that Davidson's text²⁵ has accompanying code (available from the text web site maintained by Cambridge University Press) including MATLAB scripts. In addition to these, popular antenna textbooks^{9,27–28} contain sections describing how the method of moments can be used to model specific types of antennas (typically wire antennas, though others are mentioned as well). For that reason, this section will only touch on the essentials for utilizing the MoM for antenna modeling, leaving the details for the interested reader who may choose to review these excellent textbooks and the references contained therein.

One of the most commonly used integral equations for antenna applications is the electric field integral equation (EFIE) given for metallic bodies as

$$jk_0 Z_0 \hat{n} \times \int_S \left[\mathbf{J}(\mathbf{r}') G_0(\mathbf{r}, \mathbf{r}') + \frac{1}{k_2} [\nabla'_S \cdot \mathbf{J}(\mathbf{r}')] \nabla G_0(\mathbf{r}, \mathbf{r}') \right] dS' = \hat{n} \times \mathbf{E}^{\text{inp}}(\mathbf{r}) \quad (59-10)$$

where the normal (\hat{n}) points out of a closed surface (if the surface is closed; otherwise, it points in consistent direction), $\mathbf{J}(\mathbf{r}')$ is a vector-valued unknown equivalent current, $G_0(\mathbf{r}, \mathbf{r}')$ is the free space three-dimensional Green's function while k_0 and Z_0 are the free space wave number and wave impedance, respectively. The excitation for the problem is provided by the impressed electric field, \mathbf{E}^{inp} . For an antenna, this is typically a local source such as a filamentary current, a voltage gap, or a magnetic frill current. Alternatively, it may be an impinging wave. In the former case, the antenna is operating in the transmitting mode, while for the latter case, it is operating in the receiving mode. Since most antennas are reciprocal devices, these two modes of operation are equivalent.

Solution of Eq. 59-10 exactly is not generally possible except for special cases that do not represent real-world antenna structures. Therefore a systematic approach for *numerically approximating* the solution is necessary. This is precisely what the method

of moments provides. Usually, subdomain bases (or expansion functions) are used to represent the unknown current via a collection of unknown coefficients and simple functions that are non-zero only over a portion of the computational domain (and subdomain). For example, the unknown current can be represented by

$$\mathbf{J}(\mathbf{r}') = \sum_{n=1}^N J_n \mathbf{w}_n(\mathbf{r}') \quad (59-11)$$

where the unknown, complex-valued coefficient is denoted as J_n while the generally real-valued, vector basis function is given as $\mathbf{w}_n(\mathbf{r}')$. When this is substituted into Eq. 59-10 and suitable testing (or enforcement of the boundary conditions) is performed, the resulting linear system

$$[Z_{nm}][J_n] = [f_m] \quad (59-12)$$

can be solved by some suitable method (LU decomposition, biconjugate gradient iterative solver, and so on). Note that in Eq. 59-12, the subscript n is associated with the expansion function in Eq. 59-11, while the subscript m is associated with the testing function. With these general comments in mind, it is instructive to look at two common antenna problems simulated using the MoM: wire and surface antennas.

Wire Antennas

Thin wire antennas represent the first major use of computational electromagnetics for antenna analysis and design, as well as one of the most prevalent uses even today. The widespread popularity of the Numerical Electromagnetics Code (NEC)—now in its fourth released version—is one of the reasons for widespread use. Although the basic NEC computer program is challenging to use, due to the fact that it was developed at a time when specification of an antenna using data cards was common, there are now commercial wrapper computer programs that offer a fully functional GUI and provide the input commands for NEC as its output. Examples of such front-end GUIs for NEC include: SuperNEC, MiniNEC, EZNEC, and NEC Basic, among others.

If one makes the assumption that the wire is very thin, i.e., the wire diameter is less than $\lambda_0/20$, then the following assumptions are reasonable: the surface current can be represented as a linear current and the current will only have an axial component (assume this is the z -component for convenience). This means that the thin-wire assumptions allows $\mathbf{J}(\mathbf{r}') \approx \hat{z}I(z')$. Under these conditions, the EFIE (Eq. 59-10) can be written as

$$jk_0 Z_0 \int_{-L/2}^{L/2} I(z') \left[1 + \frac{1}{k_0^2} \frac{\partial^2}{\partial^2 z} \right] G_w(z, z') dz' = E_z^{\text{inp}}(a, z) \quad (59-13)$$

where the thin-wire (a = radius) Green's function is given by

$$G_w(z - z') = \frac{e^{-jk_0 \sqrt{a^2 + (z - z')^2}}}{4\pi \sqrt{a^2 + (z - z')^2}} \quad (59-14)$$

Notice that the Green's function cannot become singular due to the approximations involved (this amounts to placing the testing points on the outer surface of the wire and the sources as filamentary currents along the axis of the wire). This particular integral equation is known as Pocklington's Integral Equation.

As an example of how a wire antenna simulation program can be used to design an antenna, a demonstration version of EZNEC (available at: www.eznec.com) was used to create the antenna pattern due to two monopoles, at approximately 300 MHz, shown in Figure 59-3. In Figure 59-3, the pattern is shown to be nearly a Cardioid pattern caused by choosing the excitation current phases and placement of the monopole radiators appropriately. The current is shown as a line with a vanishing value at the end of the monopole and a peak value at the feed (through the ground plane that is not shown but lies in the $z = 0$ plane) lying in the x - y plane as expected. In addition to currents and pattern, other engineering quantities such as the VSWR can be computed as shown for the same antenna in Figure 59-4.

Another NEC pre-processor and post-processor package is SuperNEC (<http://www.superneec.com>), which offers a 30-day demonstration copy for any user and a one-year free license for academic users. In both cases, models of up to 300 wire segments can be simulated. As an example, Figure 59-5 illustrates the currents for a Yagi-Uda antenna simulated at 300 MHz. As can be seen, the SuperNEC interface allows depiction of the current magnitude on the various wire elements. The co-polarized antenna pattern evaluated in the plane of the array (for example, the x - y plane) is shown in Figure 59-6.

Wires can also be used to simulate a surface by placing the wires sufficiently close to each other to make the wire mesh appear to be impenetrable in a manner reminiscent of a Faraday cage built using wire screens. This was a suitable and practical approach until the advent of a useful vector surface expansion function; the Rao-Wilton-Glisson (RWG) basis function was introduced in 1982.³

Surface Antennas

Many antenna structures can be represented using a surface integral equation formulation. Examples of these include: horn antennas, patch antennas, slot antennas, printed log-periodic antennas, printed spiral antennas, and Vivaldi slot antennas, among many others. Even wire antennas, especially thick wire antennas, can be modeled with a surface integral equation

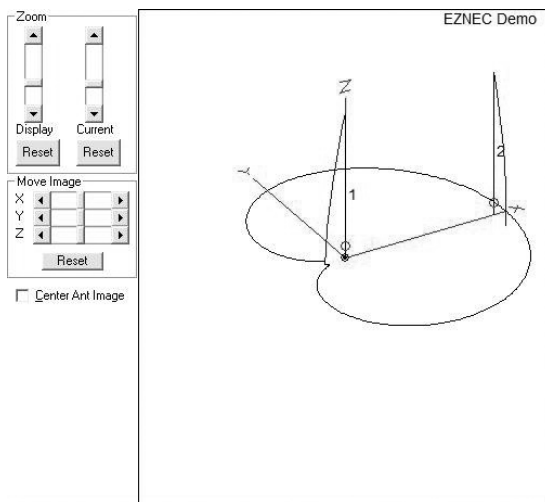


FIGURE 59-3 Radiation pattern for a two-element monopole array along with current magnitude and geometry (Calculated using EZNEC, Roy Lewallen)

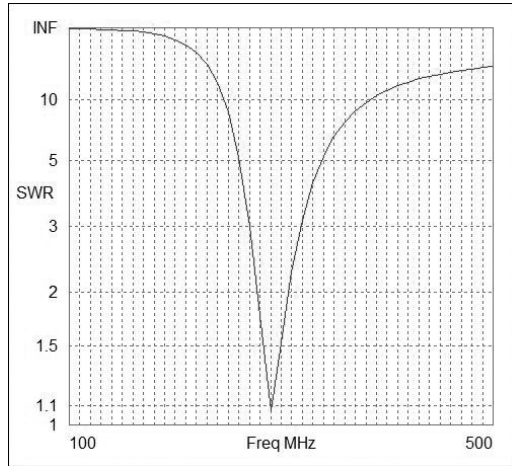


FIGURE 59-4 Standing wave ratio for the second element of a two element array (Calculated using EZNEC, Roy Lewallen)

formulation although it may be more efficient to utilize a wire formulation. Method of moments computer programs that utilize a surface integral equation differ in several ways; the most important are the Green's function used in the formulation and the basis functions used to represent the unknown quantity. For antennas in free space, the three-dimensional free space Green's function

$$G_0(\mathbf{r}, \mathbf{r}') = \frac{e^{-jk_0|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \quad (59-15)$$

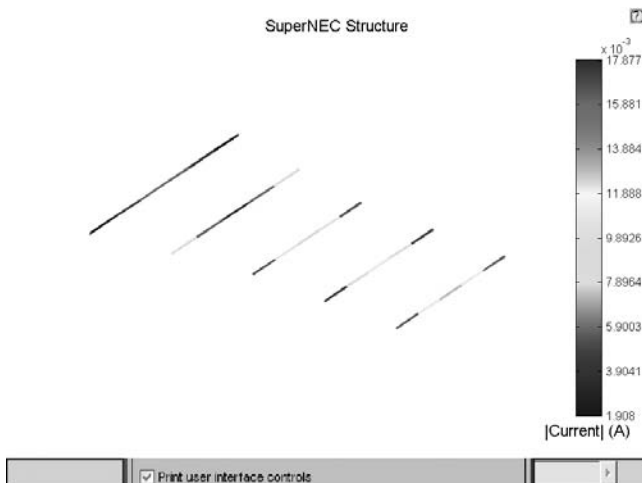


FIGURE 59-5 Magnitude of the current on a Yagi-Uda antenna (Calculated using SuperNEC)

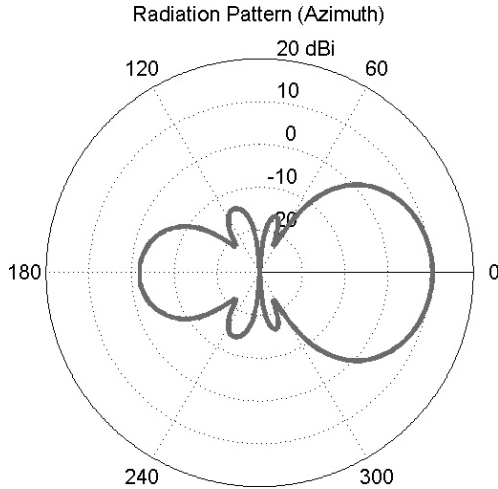


FIGURE 59-6 Co-polarized radiation pattern for a Yagi-Uda antenna (Calculated using SuperNEC)

is used. If the geometry includes a grounded material slab, then a Green's function that includes the ground plane and the material slab is generally used. In this case, the unknowns are restricted to the metallization on top of the material slab or embedded in the material slab. Hence, the order of the resulting linear system is relatively small; however, construction of the Green's function can be complex and computationally expensive unless series and/or integration acceleration techniques are used.

The RWG basis functions are probably the most widely used in the community. However, there are other surface functions that are used for particular reasons or applications. Typically a curved surface is approximated by a collection of flat facets (for example, RWG basis functions with support over flat triangles); however, in some applications, the curvature of the surface across an element is important. An example are surface elements with edge dimensions that are significant fractions of, or even larger than, a wavelength. Such elements are used to reduce the complexity of the meshing task by using higher-order expansion and surface geometry functions. Since the element dimensions are relatively large, it is important to explicitly include curvature in the element construction. This is usually accomplished by mapping a curved element to a flat element using a Jacobian.^{29–30} Note that with higher-order elements, the order of the resulting matrix is typically similar to the order of the matrix generated using low-order elements such as RWG; however, the definition of the surface can be simpler and more accurate with higher-order functions.

As an example of surface integral equation use, Makarov's book³¹ discusses the implementation of RWG basis functions and their use in solving surface antennas in free space (along with a treatment of dielectric loaded surface antennas such as patch antennas). The book comes with MATLAB scripts (these can also be found on-line at: <http://ece.wpi.edu/books/aemm/>). As an example of a complex antenna, consider the bow-tie antenna example in Chapter 8 of Makarov's book.³¹ Figure 59-7 illustrates the VSWR of that example and the current at 1 GHz. As can be seen, the antenna is relatively well matched across a wide bandwidth. The current plot indicates that the current density magnitude is greatest in the light shaded areas, for example, the antenna edges, while a dark shading indicates a region of lower current density.

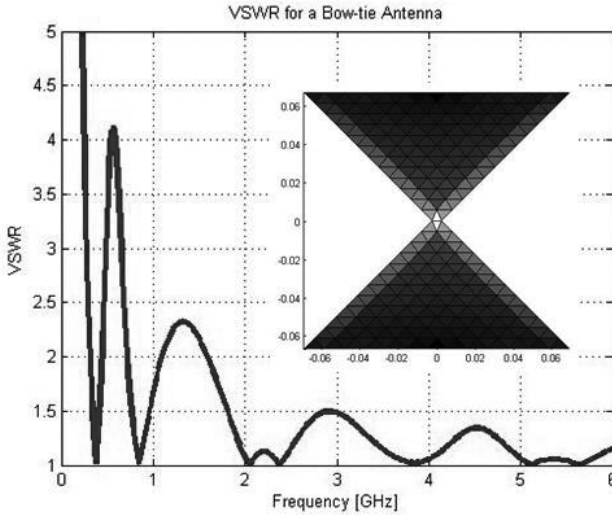


FIGURE 59-7 VSWR for a bow-tie antenna simulated using MATLAB (inset: Current density at 1 GHz; light denotes a high current density.)

Other expansion elements used in practice are low-order vector basis functions that have support over a rectangular domain.²⁹ Although rectangular basis functions limit the geometrical complexity and flexibility that can be represented without approximation, under certain circumstances, an extremely efficient solver can be utilized that significantly reduces the computational cost (and thus the time to compute). For example, if the rectangular elements are identical across the computational domain, then the resulting matrix is Toeplitz. Such matrices can be solved extremely efficiently and the computational storage requirements are much less than the general case.

As an example, consider the simulation of a patch antenna using SonnetLite, the free version of Sonnet. This antenna is printed on 30 mil Rogers Duriod/5870 and is 40mm by 30mm in dimension. The antenna is fed using a probe feed (modeled in Sonnet using an autoground port) located 16.25 mm from the left side of the antenna and centered between the top and bottom of the antenna. The local field behavior of the antenna is given in Figure 59-8, including an inset figure illustrating the current supported by the patch metal near resonance. In this, a reasonably good match is achieved over a narrow bandwidth. Improved performance can no doubt be obtained with an iterative design procedure involving adjusting the feed location. For a two-element array comprised of identical patch antennas, the S-parameters are given in Figure 59-9 (the geometry is shown as an inset figure for reference).

Modeling three-dimensional antenna geometries with a surface formulation requires either a Green's function that represents the volumetric nature of the antenna or closed surfaces that enclose the antenna volume (for example, a dielectric resonator antenna). In the former case, the resulting formulation is limited in the types of antennas that can be efficiently simulated. For the latter case, the required number of unknowns can be very large, especially if there are a large number of material interfaces. In addition, in the latter case, very thin layers can lead to a numerically ill-conditioned problem since the surface currents on different interfaces are parallel and very close to each other. Because of this, it is sometimes useful to utilize a volume integral equation³² rather than a surface integral equation.

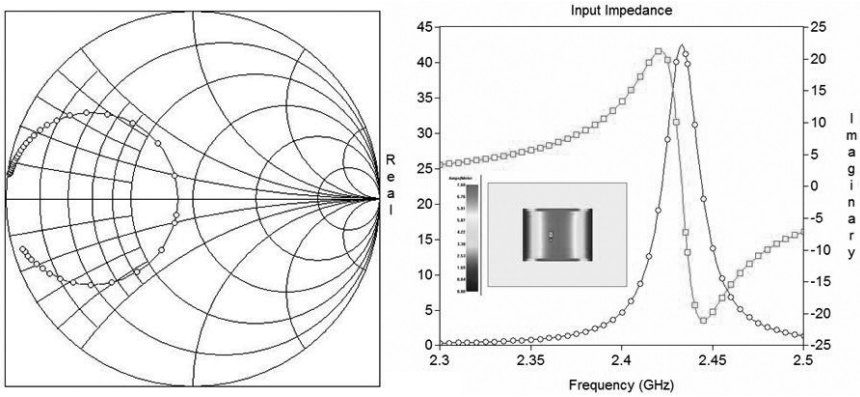


FIGURE 59-8 Performance of a single-patch antenna simulated using Sonnet (left: Smith chart, right: resistance and reactance spectrum, with inset figure illustrating resonant currents).

For example, the following integral equation³² can be used to represent dielectric materials with metallization

$$\begin{aligned} \mathbf{E}^i &= \mathbf{E} + [j\omega(\mathbf{A}_D + \mathbf{A}_M) + \nabla(\Phi_D + \Phi_M)] & \mathbf{r} \in V \\ \mathbf{E}_{\text{tan}}^i &= [j\omega(\mathbf{A}_D + \mathbf{A}_M) + \nabla(\Phi_D + \Phi_M)]_{\text{tan}} & \mathbf{r} \in S_M \end{aligned} \quad (59-16)$$

In Eq. 59-16, the volume of dielectric (V) and metal surface (S_M) are assumed to be in free space. The vector potentials (\mathbf{A}) and scalar potentials (Φ) are associated with the dielectric or metal and these are denoted by the subscript D or M , respectively. With this, the substrate need not be infinite in two dimensions and very thin dielectric layers can be simulated without unknown current surfaces that are too close. In addition, the substrate can be inhomogeneous. In contrast to most MoM implementations that utilize low-order expansion functions, higher-order functions are presented in the reference section.³³

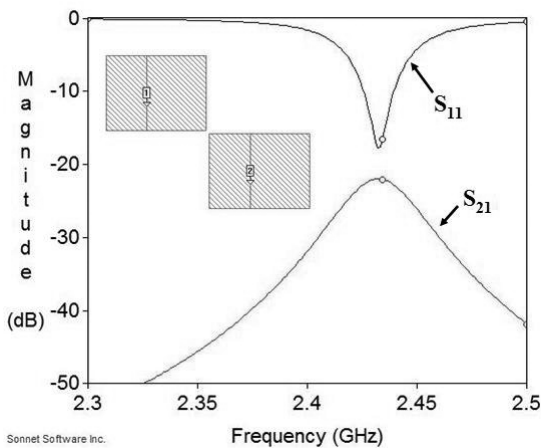


FIGURE 59-9 S-parameters for a two-element array as computed using SonnetLite (the geometry is shown as an inset figure)

Higher-order functions can be used to more accurately represent a boundary and to reduce the meshing requirements. They do not always lead to fewer unknowns since there are more degrees-of-freedom per element associated with a higher-order function as compared to a low-order function.

Finite Element (FE) Method The finite element (FE) method provides one of the most powerful formulations for antenna modeling provided that the antenna has a dielectric load, particularly if the material load is inhomogeneous or anisotropic. The FE method can be used for antennas comprised of solely conducting material; however, in general the MoM is more efficient for such radiators. The finite element method is an approach for solving Maxwell's equation by solving the vector wave equation. For time-harmonic (also known as continuous wave or frequency domain) problems, Eq. 59-2 can be recast as

$$\begin{aligned}\nabla \times \mathbf{E} &= -j\omega\mu_r\mu_0\mathbf{H} \\ \nabla \times \mathbf{H} &= j\omega\epsilon_r\epsilon_0\mathbf{E} + \mathbf{J} \\ \nabla \cdot (\epsilon_r\mathbf{E}) &= \frac{\rho_e}{\epsilon_0} \\ \nabla \cdot (\mu_r\mathbf{H}) &= 0\end{aligned}\tag{59-17}$$

where the radial frequency $\omega = 2\pi f$ is related to the frequency (f) and the surrounding space may contain an anisotropic, inhomogeneous relative permittivity (ϵ_r) and/or an anisotropic, inhomogeneous relative permeability (μ_r). However, for simplicity, isotropic materials are assumed. In Eq. 59-17, the current density (\mathbf{J}) represents sources within the computational domain. This set of first-order partial differential equations (PDE) may be re-cast as a second-order PDE, known as the vector wave equation, by taking the curl of one or the other curl expressions in (59-17) and making use of the remaining expression for simplification. So, the vector wave equation is given by

$$\begin{aligned}\nabla \times \left[\frac{1}{\mu_r} \nabla \times \mathbf{E} \right] - k_0^2 \epsilon_r \mathbf{E} &= -jk_0 Z_0 \mathbf{J} \\ \nabla \times \left[\frac{1}{\epsilon_r} \nabla \times \mathbf{H} \right] - k_0^2 \mu_r \mathbf{H} &= \nabla \times \left[\frac{\mathbf{J}}{\epsilon_r} \right]\end{aligned}\tag{59-18}$$

where the upper equation is in terms of the total electric field while the lower equation is the case when the total magnetic field is needed. In Eq. 59-18, one of the equations or the other is used depending on which of the two fields, electric or magnetic, is being sought and which of the two is most amenable for enforcement of boundary conditions. Note that when solving Eq. 59-18 for the electric field, should the magnetic field be needed in post-processing, the appropriate equation from Eq. 59-17 may be used. It is impractical to enforce these equations at every point in space. This means that a so-called weak form of the vector wave equation is typically specified, either through minimization of a functional or by application of the Rayleigh-Ritz method. For the latter, one of the wave equations (in terms of the total electric field from now on, for convenience) is tested with a vector function. In this, the dot-product of the vector function is taken and the resulting scalar expression is integrated throughout the domain of the test function. As a result, Eq. 59-18 in the weak form is written after some manipulation (such as use of the First Vector Green's Theorem³⁴ to transfer a curl operator off of the unknown field onto the testing function, \mathbf{W}_i) as shown in

$$\int_V \left[\frac{\nabla \times \mathbf{W}_i \cdot \nabla \times \mathbf{E}}{\mu_r} - k_0^2 \epsilon_r \mathbf{W}_i \cdot \mathbf{E} \right] dV - jk_0 Z_0 \oint_S \mathbf{W}_i \cdot (\hat{n} \times \mathbf{H}) dS = -jk_0 Z_0 \int_V \mathbf{W}_i \cdot \mathbf{J} dV\tag{59-19}$$

where the subscript i refers to the testing function. Several observations can be made regarding this weak-form equation. In Eq. 59-19, the surface magnetic field will be provided either by a local condition or by an integral relation. In either case, the surface magnetic field is proportional to the surface electric field. It's a given for this expression that isotropic materials have been assumed, and the resulting matrix developed from Eq. 59-19 will be symmetric (since the electric field and testing functions may be interchanged without changing the form of the expression). In addition, the surface term in Eq. 59-19 allows the introduction of boundary conditions on $S = \partial V$. Through these boundary conditions, an external source may be included such as an impinging plane wave. A series of relatively easy-to-read papers for FEM beginners have been featured in the EM Programmer's Column of the IEEE Antennas and Propagation Magazine in recent years.³⁵⁻³⁶ The reader is encouraged to read such papers and their references for a greater understanding of the FE method.

Most implementations of the finite element method for electromagnetics primarily differ in the elements used to expand the unknown field (and usually the testing functions as well since these are similar) and the method to enforce the boundary conditions. The most popular elements for solution of the finite element equations are the brick, right prism, and tetrahedron. These are shown in Figure 59-10 with the brick on the left, the prism in the center, and the tetrahedron on the right.

By far the most popular, due to its flexibility, is the tetrahedron. Nevertheless, both the brick and right prism have their uses in that they are particularly easy to implement and, most importantly, rather easy to generate appropriate meshes for. Usually, the expansion functions are lowest mixed-order edge elements (for example, the field representation is constant for component of the field parallel to the edge and linear for the component of the field normal to the edge). However, higher-order expansion functions³⁷ and hierarchical functions³⁸ have been used by a number of researchers. The user will find that the major discriminator between finite element codes is not the element shape necessarily, nor even the order of the expansion; rather, it is the inclusion of a powerful mesh generator. It is not an understatement to say that a large portion, and sometimes the majority, of effort expended by antenna engineers in using a finite element code is in mesh generation rather than data collection.

The second major discriminator is the mesh termination condition. The following three methods are most popular:

- Absorbing boundary conditions³⁹
- Perfectly matched layers⁴⁰
- A boundary integral⁴¹

The first method, absorbing boundary conditions (ABCs), is widely used since it is particularly easy to implement without having to use user-defined parameters that

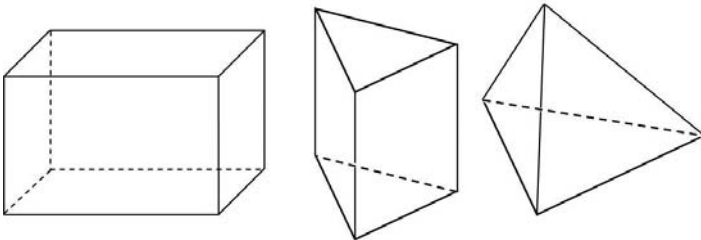


FIGURE 59-10 Typical finite element shapes: brick (left), right prism (center), and tetrahedron (right)

impact performance. The cost of using these ABCs is usually a larger computational domain since the mesh boundary must be maintained some distance from the structure being implemented. The distance required is dependent on the particular ABC and on the problem. In general, antenna simulations require a larger stand-off distance than scattering problems due to the need for accurate near-field quantities such as driving-point impedance. Perfectly matched layers (PMLs) are another local (or in other words, dependent only on field behavior in the vicinity of the PML) mesh truncation technique that generally requires less stand-off distance than ABCs. However, there are user-defined (or at least programmer-defined) parameters and typically, the condition number of the resulting matrix is poorer as compared to the ABC. The final method, a boundary integral (BI), is a global rather than local condition. It requires no stand-off distance; however, the resulting matrix is partially full due to the BI, and as a result, the demand for memory and solution time can be quite high. In addition, other aspects of the method formulation and implementation can impact solution performance. For example, the use of adaptive meshing⁴² is a common feature in commercial programs these days that largely reduces the need for specialized meshing expertise so that the essential field behavior is captured.

As an example of how the finite element method can be used for antenna analysis and design, consider a 40mm by 30mm cavity-backed antenna placed in a cavity that is 80 mm by 60 mm. The cavity is filled with 30 mil thick Rogers Duriod/5870. This particular antenna will be analyzed using the computer program provided on the author's web page (to obtain a copy of the source code, see the URL cited in the next section). This program utilizes right prism elements⁴³ and is written in MATLAB with a functional graphical user interface (GUI). The program is initiated in MATLAB by running the msGUI.m script. The cavity is divided into 32×24 cells that are then subdivided into triangles. The program then extrudes these triangles into right prism elements. Since the substrate is electrically thin, one layer of elements is sufficient to resolve the field behavior along the normal direction. The surface mesh, Smith Chart, and normalized electric field within the cavity are shown in Figure 59-11.

The fields shown are the normal component of the electric field. The patch antenna is excited using a probe-feed from the base of the cavity to the patch slightly left of the antenna center along the top-to-bottom centerline. This method of feeding excites the fundamental mode for this antenna. Notice that the normal fields are zero along the centerline of the antenna while those fields are a peak near the edges of the patch. This field distribution can be represented using a cosine function. The phase of the field distribution left of the centerline as compared to right of the centerline is 180 degrees out-of-phase resulting in radiation.

This particular example does not utilize the most significant feature of a finite element formulation: the ability to model inhomogeneous antenna loading materials. Simulation of grounded, homogeneous dielectrically-loaded antennas is most efficiently analyzed using an integral equation method (see the previous example using Sonnet). However, if the antenna has inhomogeneous (or anisotropic) materials in its construction, the finite element method is often the most efficient and accurate method to use for simulating resonant structures

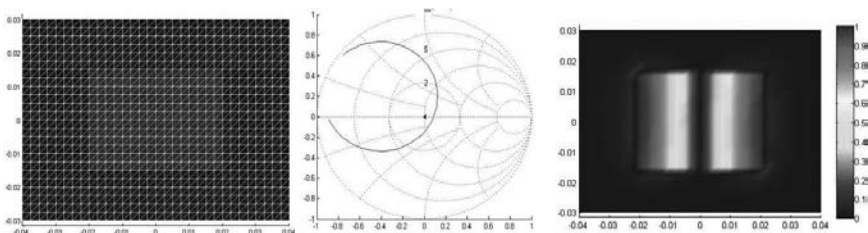


FIGURE 59-11 Surface mesh (left), Smith chart (center), and normal field distribution (right) for a cavity-backed patch antenna simulated using the FE-BI method

(for example, the class of antennas that is best simulated using a frequency-domain solver). This is also true for the case of cavity-backed antennas, when the cavity is odd-shaped regardless of the cavity-fill material.

One comment is in order concerning simulation of realistic antennas in the MATLAB environment. Unless considerable effort is undertaken to vectorize (express the mathematical manipulations in terms of data vectors and matrices), parallelize, or compile as executable code, MATLAB is not an efficient platform for computing *large* problems. The definition of large is in the “eye of the beholder,” in the sense that for a given algorithm and the time a user is willing to wait for an answer, the computational size that constitutes a large problem varies. Nevertheless, for understanding the underlying concepts of various computational methods, it is an effective platform for discovery. It is particularly useful for students due to their general familiarity with MATLAB.

Other Frequency-Domain Methods

The method of moments and finite element methods are arguably the most prevalent frequency-domain CEM solvers in use today. Nevertheless, there are excellent alternative methods that have very specific advantages in certain circumstances and other methods that are growing in popularity and impact. One of the most effective means for computing the radiation by antennas in the presence of very large structures involves the use of high frequency (asymptotic) approximations such as the Uniform Theory of Diffraction (UTD). This method was extensively investigated by various research groups around the world during the 1970s and beyond, most notably by the group at The Ohio State University, along with their colleagues from around the world who spent time at the ElectroScience Laboratory as students, post-doctoral researchers, and as visiting scientists.⁴⁴⁻⁴⁵

Another method of considerable note is the finite integration technique (FIT) mentioned earlier (in its time-domain incarnation). FIT is formulated using the integral form of Maxwell’s equations, either in the time- or frequency-domains. FIT can also be used in conjunction with a high frequency method such as UTD.⁴⁶ The method of lines is another method for simulating electromagnetic phenomena and it has been used to simulate conformal antennas.⁴⁷ A good reference for various conformal antenna formulations is provided by a recent text.⁴⁸

Another approximate technique for simulation of antennas installed in a large structure involves the use of reciprocity.⁴⁹ In this situation, the antenna radiating currents are computed and then integrated with the currents due to the structure resulting in the radiation pattern of the antenna in the presence of the structure. Although the paper is used for circular cylindrical structures, it can be used for more general shapes. These hybrid formulations are very powerful, and at the dawn of the 21st century, the only practical method for solving installed antenna design problems when the structure is electrically very large. It is probable that through advances in computer technology and in full-wave solution methods (for example, fast multipole methods among others) asymptotic methods will be needed less and less. Having said that, if history teaches us anything, design problems will become more complex and larger over time and thus, asymptotic and hybrid methods may be needed long into the future.

59.4 SUMMARY OF PUBLIC DOMAIN AND COMMERCIAL CODES: CAVEAT EMPTOR

Computational electromagnetics has undergone a typical evolution during the past half century: What was merely a research topic in the 1960s through the 1980s is now a thriving industry. Accordingly, many CEM tools are available for antenna design. Some of these

are in the public domain (some even provide the source code so that users can modify the program for their own needs), while others are commercial offerings. As the section title indicates, the user (or buyer) must beware in assuming a particular program will solve their problem, or even be correct, either in formation, assumption, or coding. In general, free programs have absolutely no guarantee as to their accuracy while commercial codes are generally easier to use and typically have technical professionals (such as application engineers) to help adopters generate solutions. One prevailing dictum of CEM users should be: “*Purgamentum init, exit purgamentum*”! Inexperienced users of any CEM computer program are cautioned to validate both the program and their use of the program by comparing their simulation with reference data. It is better to compare a user-generated case rather than running examples provided by the program developers since in that manner, the user will not only be testing the program, but their interpretation of how to give instructions to the program. By all means, initial users should explore the examples that are now standard for CEM codes; but at some point, users must try their own hand at developing a model before they try to solve the design challenge that originally prompted their need (or at least desire) for the CEM code. A brief listing of public domain and commercial codes is given next. In no manner does this constitute a comprehensive list and over the lifetime of this handbook, it can be expected that Internet links will be broken and that companies will no longer offer these codes. Hence, “Let the buyer beware!!!”

Disclaimer: *This section is not meant in any way to be an endorsement of any methods, computer programs, or company's offerings. It is merely meant to be an information source for antenna designers interested in using powerful CEM design tools. This is not a comprehensive list by any means and any omissions in the interest of brevity are regrettable, but not intentional.*

Public Domain Codes

Many computer programs are offered free-of-charge by the developers, in the interest of promoting the adoption of the technology or as a means of technology transfer. This was particularly true in the early days of CEM; it is less so in the 21st century as the intellectual property value of such codes is better understood. Nevertheless, very powerful computer programs may be downloaded, and in most cases, modified by the user since source code is available. Codes may be downloaded from

- **www.egr.msu.edu/~kempel/HandbookCodes** This is where you can find the MATLAB scripts used in this chapter. The FD-TD code is provided by Mr. Steven Cossmann and the FE-BI code by Mr. Lanwu Zhao (developed with partial support from the National Science Foundation under grant DUE-0231312).
- **<http://www.emclab.umn.edu/codes.html>** The University of Missouri at Rolla (UMR) maintains an excellent listing of public domain codes available on the Internet. The reader is encouraged to investigate this site when looking for a free program, especially the finite element code—EMAP—developed by the students and faculty at UMR.

Antenna Textbooks with Codes

Many antenna textbooks now come with accompanying computer programs to aid the student and engineer. Sometimes these codes come with a CD-ROM disk, while in some cases a related web site is provided for users to download code. This latter method has the distinct advantage in that program updates may be provided in a convenient manner.

Unfortunately, it has the disadvantage in that the site must be maintained and in practice, links to that web site may break over time. Nevertheless, the software provided with modern antenna textbooks is very useful for users, both to learn how antennas function and to provide validation data in the context of accurate CEM codes. Some texts have been discussed previously; however, for the convenience of the reader, they are repeated in the following list. Note that there are undoubtedly many more antenna textbooks with complementary computer programs; brevity necessitates an unfortunately incomplete list:

- David Davidson from the University of Stellenbosch in South Africa has written a textbook²⁵ describing various CEM techniques. A major application of that program is antenna design and performance analysis. Software for the text is provided via a resources web site maintained by Cambridge University Press (<http://www.cambridge.org>). Both MATLAB scripts and FEKO files are provided that support the material in the book. Note that Appendix F lists additional resources found on the Internet.
- Sergey Makarov of Worcester Polytechnic Institute wrote a textbook²⁶ that describes the modeling of various antenna geometries (such as patches, slots, inverted-F, and so on) using the method of moments. A new web site (<http://ece.wpi.edu/mom>) compared to the one cited earlier in the chapter provides MATLAB scripts and a user manual for using the scripts or for modifying them as needed for a different type of antenna. A companion paper illustrates some of the capability of this design tool.³² This package was developed with partial support from the National Science Foundation under grant DUE-0231312.
- John D. Kraus and Ron Marhefka from Ohio State University, in the third edition of their popular antenna textbook,²⁷ provided a web site for several antenna design computer programs (<http://www.antennas3.com>). These accurate and efficient programs are useful for understanding antenna function as well as providing a method for validation.
- Constantine Balanis from Arizona State University, in the third edition of his well-regarded antenna textbook,²⁸ provides various antenna modeling codes in both MATLAB and Fortran (from the second edition) via a CD-ROM. The scripts provided are especially useful for students and for antenna design engineers interested in validating a CEM tool.
- Atef Elsherbeni and Matthew Inman from the University of Mississippi have developed a textbook⁵⁰ that tightly integrates visualization of antenna properties (such as pattern impedance, and so on). Along with the book, the authors have developed software (www.adv-program.com) that implements standard formula for various wire-type antennas (such as dipoles, helices, and monopoles) in both single element and array configurations. The program has a GUI and can provide reference data for other programs that implement CEM techniques.

Commercial Codes

Commercial computer programs are not only powerful, but also designed to be easy-to-use. That last term is something that is in the “eye-of-the-beholder.” What software designers consider easy-to-use and dummy-proof often does not correspond to the expectations of the user. Nevertheless, commercial CEM codes are typically better designed from the user point-of-view than public domain codes. They typically have a Graphical User Interface (GUI) and application engineers that can (for a fee) help users obtain the best possible calculations for a given problem. During the 1990s and 2000s, the number and diversity of commercial CEM offerings have expanded greatly. The following list describes some of the most popular commercial CEM codes. No doubt during the lifetime of this edition, the offerings will change considerably:

- **www.ansoft.com** The Ansoft Corporation offers a variety of CEM solutions for antenna design, electromagnetic compatibility, and other applications. The computer codes offered by Ansoft utilize the finite element method, the method of moments, and circuit simulators. These computer programs are some of the most powerful in the industry and are widely used.
- **www.cst.com** Computer Simulation Technology offers widely used CEM tools based on the finite integration technique (FIT) and the method of moments. Both time-domain and frequency-domain solvers are available.
- **www.emssuusa.com** Electromagnetic Software and Systems offers a hybrid method of moments, finite element, and high frequency (uniform theory of diffraction) computer code (also known as FEKO). This tool is widely used for antenna design and for EMC analysis in the automobile industry.
- **www.remcom.com** Remcom offers a finite difference-time domain method computer program that can be used for antenna design, scattering analysis, and specific absorption rate (SAR) calculations, among other uses. It comes with a well-developed GUI and support for distributed parallel computers (for example, clusters).
- **www.sonnetusa.com** Sonnet Software offers a suite of full-wave electromagnetic design tools based on a frequency-domain method of moments for planar geometries. It provides a powerful GUI and allows for separation between the design workstation and the computational workstation. A free reduced-capability version is available for download.
- **www.wipl-d.com** WIPL-D is a powerful antenna design program capable of modeling antennas (and other structures) comprised of both metal and dielectric materials. WIPL-D is becoming increasingly popular amongst antenna engineers due to its flexibility and accurate solutions.
- **www.comsol.com** Comsol Software is based in Sweden with offices worldwide. Their main product is FEMLAB, a finite element-based simulation tool. Their claim to fame is the ability to include multiple physical and chemical models in the simulation. Thus, simultaneous simulation of both thermal and electromagnetic properties is facilitated by FEMLAB. It has a powerful mesh generator and various pre- and post-processing capabilities.
- **www.zealand.com** Zealand offers the popular integral equation-based solver, IE3D. This program is widely used for simulating planar, microstrip structures such as patch antennas with feed networks.

59.5 SOURCES FOR PARALLEL PROGRAMMING INFORMATION

Computer architectures have undergone considerable, and truly impressive, changes over the past 30 years. In the 1970s, computers for engineering design were primarily housed in centralized computing centers and used in a time-share manner (often with punch cards as the input method of choice). In the 1980s, with the advent of the personal computer, it became possible to put a computer on the desk of every engineer. This led to more detailed antenna designs, especially when the Numerical Electromagnetic Code (NEC) was released by Lawrence Livermore National Laboratory. In the 1990s, the widespread use of powerful engineering workstations led to increasing use of powerful computational electromagnetics tools in antenna design. During that decade, the workstation and personal computer merged (or more accurately, the personal computer became a workstation)

leading to even more widespread use of CEM codes. Simultaneously, during the latter 1980s and early 1990s, vector-based computing became available through CRAY Computers, leading to changes in computational algorithms that specifically made use of vector processors. Once the practical limit of vector processors have been reached (in terms of computations per second versus cost), the next major change in computation involved the use of many processing units simultaneously. These parallel processing computers offered the promise of nearly unlimited computational power as more and more “brains” were brought to bear on a problem. In reality, the community quickly learned that parallel programming is not trivial and that very real limits exist on the scalability of many CEM approaches. Nevertheless, for those fortunate souls with access to massively parallel computers, these new architectures offered unprecedented computing power.

Parallel computers traditionally are symmetric multiprocessing (SMP) or shared memory architectures. That means that each processor has access to all of the memory in the system. SMP machines are particularly easy to program since the processors and memory are tightly connected together. However, large SMP machines are highly-engineered and therefore rather expensive. Therefore, a popular parallel computer architecture available in the late 1990s and early 21st century is the so-called Beowulf cluster. These computers are made from a number (sometimes a large number) of commodity compute nodes connected together with some network fabric. The original network fabrics were 10 or 100 MHz Ethernet connections (identical to those used to form a local area network). Currently, the most commonly used network fabric is a Gigabit Ethernet fabric. It is becoming increasingly common to have higher speed (or lower latency) fabrics such as Myranet or Infiniband as the acquisition cost of these technologies is reduced. Clusters are distributed memory architectures that often have a lower acquisition cost as compared to a comparable number of processors in an SMP system. However, the programming requirements for a distributed cluster are often more significant than for an SMP system.

Recently, SMP machines are seeing a renaissance through the introduction of multiple core processors. Traditionally, each processor had a single compute core. Moore’s Law dictated a doubling of transistors every 18 months and as a result of this, it dictated an increase in computing power. However, it became evident as the industry entered into the 21st century that such an expectation could not be sustained indefinitely. One way to increase computing performance was to fabricate two (or dual) cores on each processor. This way every single processor computer became an SMP while traditional thin (minimal capability) nodes for a cluster went from having two processors (a so-called 2-way node) to four cores (a so-called 4-way node). Most laptop computers offered for sale in 2006 now have dual-cores and so are properly SMP computers. Realistically, with only two cores, efficient parallel programming is not practical due to overhead requirements. Nevertheless, the trend is clear. SMP computing is here to stay and will become ubiquitous. Indeed, processors with four cores (e.g., quad-core) were released at the end of 2006 to original equipment manufacturers (OEMs) and to the public. This means that computers will become 4-way SMP machines while thin cluster nodes will become 8-way SMP nodes.

It is therefore important for CEM code developers to become skilled at utilizing the dominant parallelization paradigms. It is also important for consumers of either public domain or commercial CEM codes to understand the difference between the two types of parallel architectures (SMP vs. distributed cluster) and the parallelization techniques used by those computer programs. The two dominant parallel programming models in use today are OpenMP and MPI. These two protocols are used to instruct the parallel computer on how to distribute the workload when more than one thread (akin to a core or processor) is being used simultaneously to solve a particular computational challenge. Each of these application programming interfaces (API) will be discussed, briefly, in turn. Finally, some brief comments on the use of the MATLAB distributed toolbox will be made along with the program environment, STAR-P.

OpenMP

OpenMP is perhaps the simplest way to parallelize a program for use on an SMP computer. Recall that an SMP computer has a number of threads within a single image (essentially, within a single node). The other major paradigm for parallelization of a program is Message Passing Interface (MPI). In MPI, the sending and receiving of data between nodes must be explicitly stated, as it is much more complex to implement than OpenMP. However, MPI can be used effectively on a cluster system while OpenMP cannot. With the emergence of quad-core processors, a four-processor computer has 16 cores and is in reality a 16-way system. This is could be the optimal computer for CEM codes utilizing a matrix solver.

There are a large number of OpenMP tutorials on the Internet⁵¹⁻⁵³ and textbooks including material on OpenMP.⁵⁴⁻⁵⁶ The examples in these texts are often in C (or C++); however, for those who still prefer Fortran, there are a large number of code examples in Fortran 77 and Fortran 90. These are the predominant languages used for scientific and engineering computing.

Parallelization of a program, at the most basic level, requires only a few commands describing how to partition the work amongst a number of threads for loop iteration. The beginning of a parallel region must be declared as does the end of that parallel region. Within the parallel region, loops that the programmer wishes to parallelize are identified individually along with some descriptions on how to parallelize the loop. Data within a parallel region is either “shared” or “private” with the assumed state varying between implementations of OpenMP. Shared data is available to all threads while private data is individual to a given thread. It is important to understand how the data is identified, especially when reduction of data strings occurs (for example, addition of data vectors from various threads into a single data vector in the master thread).

In general, users of OpenMP parallelized computer programs need to do little to nothing if the program is written in a user-friendly manner. One environment variable that may need to be set is the number of available threads. In OpenMP, the command varies a bit based on the operating system and, for either Unix or Linux, the shell being used. For example, in the Linux bash shell, the command `export OMP_NUM_THREADS=#` where “#” is the number of threads is used. Programs then utilize that environment variable information during execution. Typically this number should be set less than or equal to the number of available cores.

It is generally best to have a parallel region as large as possible to minimize overhead and to provide for the best synchronization of the program threads. This is due to the fact that there is an implied barrier as the end of parallel region is reached. What this means is that when an end of the parallel region is reached, all threads are synchronized. This means that the slowest thread to be executed defines when the program execution outside the parallel region can commence. Load balancing then becomes an important consideration for how well the program executes on a parallel machine. In a similar light, use of the parallel region construct will generally be favored over loop-level parallelization; however, it comes with the cost of greater programming complexity (to ensure that all the code in the parallel region is compatible with parallel execution).

MPI

Message Passing Interface (MPI) is the second predominant parallelization paradigm in use at the dawn of the 21st century. It has the distinct advantage (compared to OpenMP) of being useful on *distributed memory* computers, commonly referred to as clusters. MPI requires considerably more investment by the programmer in terms of effort, since with MPI the programmer must tell the program how to share information between nodes

(these details are hidden from the user with OpenMP). For example, if a matrix-vector product is to be parallelized using MPI, the data for a given worker must be sent by the master (including the type of data and number of entries) where the master identifies which worker will receive the data. Likewise, the worker must explicitly receive the data (knowing the format, data type, and so on), do the work needed, and then send the resulting data back to the master. In this case, the master must receive the data knowing which worker sent a particular set of data.

MPI is very powerful in that it can be used on both SMP and cluster computers. In general, the performance of a parallel program using MPI is greater than that of OpenMP, especially in terms of scalability. Loosely speaking, scalability indicates the number of threads that can be efficiently used. Many CEM algorithms by their nature do not scale beyond 8-16 threads except for some hardware implementations (for example, the differences in memory access methods, interconnects, and so on, will impact scalability). One reason for this is the large amount of memory required for the method as compared to the relatively narrow data pipes between the compute cores and the main memory. Note that FD-TD is the CEM method that is well-known to scale beyond 16 threads with even moderate levels of parallelization skill. There are many textbooks that discuss MPI programming.^{55,57-58}

For the user, the level of effort for running (rather than parallelizing) a program is not excessive; however, it is greater than what is needed for OpenMP. Minimally, the program must be compiled and linked with a particular implementation of the MPI libraries. These libraries are specific to the MPI implementation and to the interconnect fabric. Hence, a program linked to the MPI library for Myranet will not operate properly on a cluster that uses Infiniband. To run a program that has been parallelized with MPI, rather than running the executable code directly as one does with most Unix/Linux programs, the user must invoke the MPI execution wrapper. For example, to run the program foo, many MPI implementations require the following command for execution: `mpirun foo`. The user needs to consult the relevant documentation for both the cluster and the program in order to determine all the necessary environment variables and execution scripts.

MATLAB and STAR-P

In addition to traditional high-level languages, such as C, C++, and Fortran 90, a new type of parallel environment is being distributed and continuously improved. MATLAB, as stated previously in this chapter, is widely used in the engineering community. MATLAB offers a Distributed Computing Toolbox (see <http://www.mathworks.com>) that simplifies the implementation of MATLAB scripts for parallel execution on cluster computers. The user should be cautioned that placing this toolbox on a cluster system is not as straightforward as serial toolboxes since the distributed toolbox operation must be compatible with the scheduler and operating system implementation. Note that if the program is not compiled, execution of large problems will not usually be very efficient due to the interpretive nature of the MATLAB programming environment.

STAR-P is another recent product that simplifies the development and utilization of parallel programs on distributed computers. STAR-P is used in conjunction with interactive desktop tools, such as MATLAB. It acts as a bridge between MATLAB scripts and a cluster (or SMP) computer. So, its function is related to the Distributed Computing Toolbox offered by MathWorks. More information concerning STAR-P can be found at: <http://www.interactivesupercomputing.com>. The particular solution for any given application and user is dependent on the cluster system implementation. Nevertheless, it is prudent to test particular applications with the various options such as the Distributed Computing Toolbox, STAR-P, or other computing environments before promising results to customers!

REFERENCES

1. K. S. Yee, "Numerical Solution of Initial Boundary Problems Involving Maxwell's Equations in Isotropic Media," *IEEE Trans. Ant. Propagat.*, 14, (May 1966): 302–307.
2. G. Burke and A. Poggio, "The Numerical Electromagnetic Code (NEC)," LLNL Technical Report UCID-18834 (November 1980).
3. S. M. Rao, D. R. Wilton, and A. W. Glisson, "Electromagnetic Scattering by Surfaces of Arbitrary Shape," *IEEE Trans. Ant. Propagat.*, vol. 30 (May 1982): 409–418.
4. J. C. Nédélec, "Mixed Finite Elements in R³," *Numer. Math.*, vol. 35 (1980): 315–341.
5. J. M. Jin and J. L. Volakis, "A Finite Element Boundary-Integral Formulation for Scattering by Three-Dimensional Cavity-Backed Apertures," *IEEE Trans. Ant. Propagat.*, vol. 39 (January 1991): 97–104.
6. W. C. Chew, J. M. Jin, E. Michielssen, and J. M. Song (eds.), *Fast and Efficient Algorithms in Computational Electromagnetics* (Boston: Artech House, 2001).
7. E. Bleszynski, M. Bleszynski, and T. Jaroszewicz, "AIM: Adaptive Integral Method for Solving Large-Scale Electromagnetic Scattering and Radiation Problems," *Radio Science*, vol. 31 (1996): 1225–1251.
8. T. Weiland, "A Discretization Method for the Solution of Maxwell's Equations for Six-Component Fields," *Electronics and Communications AEUE*, vol. 31 (1977): 116–120.
9. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd Ed. (New York: Wiley & Sons, 1998).
10. A. Taflové and S. C. Hagness, *Computational Electromagnetics—The Finite-Difference Time-Domain Method*, 3rd Ed. (Boston: Artech House, 2005).
11. K. Kuntz and R. Luebbers, *The Finite Difference Time Domain for Electromagnetics* (Boca Raton: CRC Press, 1993).
12. E. K. Miller, A. J. Poggio, and G. J. Burke, "An Integro-Differential Equation Technique for the Time-Domain Analysis of Thin Wire Structures," *J. Comp. Physics*, vol. 12 (May 1973): 24–48.
13. W. C. Chew, J. M. Jin, E. Michielssen, and J. M. Song (eds.), *Fast and Efficient Algorithms in Computational Electromagnetics* (Boston: Artech House, 2001).
14. D. Poljak and C. Y. Tham, "Integral Equation Techniques in Transient Electromagnetics," *Advances in Electrical and Electronic Engineering*, 3rd Ed. (Southampton: WIT Press, 2003).
15. D. S. Weile, G. Pisharody, N.-W. Chen, B. Shanker, and E. Michielssen, "A Novel Scheme for the Solution of the Time-Domain Integral Equations of Electromagnetics," *IEEE Trans. Antennas and Propagat.*, vol. 52 (January 2004): 283–295.
16. K. Aygün, B. C. Fisher, J. Meng, B. Shanker, and E. Michielssen, "A Fast Hybrid Field-Circuit Simulator for Transient Analysis of Microwave Circuits," *IEEE Trans. Microwave Theory Tech.*, vol. 52 (February 2004): 573–583.
17. R. Holland, V. P. Cable, and L. C. Wilson, "Finite-Volume Time-Domain (FVTD) Techniques for EM Scattering," *IEEE Trans. Electromagn. Compat.*, vol. 33 (1991): 281–294.
18. J.-F. Lee, R. Lee, and A. Cangellaris, "Time-Domain Finite-Element Methods," *IEEE Trans. Ant. Propagat.*, vol. 45 (March 1997): 430–442.
19. A. Monorchio, A. R. Bretones, R. Mittra, G. Manara, and R. G. Martin, "A Hybrid Time-Domain Technique that Combines the Finite Element, Finite Difference, and Method of Moments Techniques to Solve Complex Electromagnetic Problems," *IEEE Trans. Ant. Propagat.*, vol. 52 (October 2004): 2666–2674.
20. D. J. Riley, J.-M. Jin, Z. Lou, and L. E. R. Petersson, "Total- and Scattered-Field Decomposition Technique for the Finite-Element Time-Domain Method," *IEEE Trans. Ant. Propagat.*, vol. 54 (January 2006): 35–41.
21. R. Lee, "A Note on Mass Lumping in the Finite Element Time Domain Method," *IEEE Trans. Ant. Propagat.*, vol. 54 (February 2006): 760–762.
22. R. Ehmann, B. Wagner, and T. Weiland, "Farfield Calculations for Car Antennas at Different Locations," *IEEE Trans. Magnetics*, vol. 33 (March 1997): 1508–1511.

23. P. R. Rousseau and P. H. Pathak, "Time-Domain Uniform Geometrical Theory of Diffraction for a Curved Wedge," *IEEE Trans. Ant. Propagat.*, vol. 43 (December 1995): 1375–1382.
24. R. F. Harrington, *Field Computation by Moment Methods* (New York: IEEE Press-Wiley, 1993).
25. D. B. Davidson, *Computational Electromagnetics for RF and Microwave Engineering* (Cambridge, Mass: Cambridge University Press, 2005).
26. S. N. Makarov, *Antenna and EM Modeling with MATLAB* (New York: Wiley & Sons, 2002).
27. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd Ed. (Boston: McGraw-Hill, 2002).
28. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd Ed. (New York: Wiley & Sons, 2005).
29. A. F. Peterson, S. L. Ray, and R. Mittra, *Computational Methods for Electromagnetics* (Piscataway, NJ: Wiley-IEEE Press, 1997).
30. A. F. Peterson, *Mapped Vector Basis Functions for Electromagnetic Integral Equations* (San Rafael: Morgan & Claypool, 2006).
31. S. N. Makarov, *Antenna and EM Modeling with Matlab* (New York: Wiley, 2002).
32. S. N. Makarov, S. D. Kulkarni, A. G. Marut, and L. C. Kempel, "Method of Moments Solution for a Printed Patch/Slot Antenna on a Thin Finite Dielectric Substrate Using the Volume Integral Equation," *IEEE Trans. Ant. Propagat.*, vol. 54 (April 2006): 1174–1184.
33. M. Djordjevic and B. M. Notaros, "Double Higher Order Method of Moments for Surface Integral Equation Modeling of Metallic and Dielectric Antennas and Scatterers," *IEEE Trans. Ant. Propagat.*, vol. 52, (August 2004): 2118–2129.
34. C-T. Tai, *Dyadic Green's Functions in Electromagnetic Theory*, 2nd Ed. (Piscataway, NJ: IEEE Press, 1994).
35. D. B. Davidson, "Implementation Issues for Three-Dimensional Vector FEM Programs, *IEEE Antennas Propagat. Mag.*, vol. 42 (December 2000): 100–107.
36. A. Awadhiya, P. Barba, and L. Kempel, "Finite-Element Method Programming Made Easy???", *IEEE Antennas Propagat. Mag.*, vol. 45, (August 2003): 73–79.
37. R. D. Graglia, D. R. Wilton, and A. F. Peterson, "Higher Order Interpolatory Vector Bases for Computational Electromagnetics," *IEEE Trans. Antennas Propagat.*, vol. 45 (March 1997): 329–342.
38. L. S. Andersen and J. L. Volakis, "Accurate and Efficient Simulation of Antennas Using Hierarchical Mixed-Order Tangential Vector Finite Elements for Tetrahedral," *IEEE Trans. Antennas Propagat.*, vol. 47 (August 1999): 1240–1243.
39. J. L. Volakis, A. Chatterjee, and L. C. Kempel, *Finite Element Method for Electromagnetics* (Piscataway, NJ: IEEE Press, 1998).
40. Z. S. Sacks, D. M. Kingsland, R. Lee, and J-F Lee, "A Perfectly Matched Anisotropic Absorber for Use as an Absorbing Boundary Condition," *IEEE Trans. Antennas Propagat.*, vol. 43 (December 1995): 1460–1463.
41. J-M. Jin, *The Finite Element Method in Electromagnetics*, 2nd Ed. (New York: Wiley, 2002).
42. M. Salazar-Palma, T. K. Sarkar, L-E. Garcia-Costillo, and T. Roy, *Iterative and Self-Adaptive Finite-Elements in Electromagnetic Modeling* (Boston: Artech House, 1998).
43. L. C. Kempel, "Implementation of Various Hybrid Finite Element-Boundary Integral Methods: Bricks, Prisms, and Tets," *1999 ACES Meeting*, Monterey, CA (1999): 242–249.
44. D. A. McNamara, *Introduction to the Uniform Theory of Diffraction* (Boston: Artech House, 1990).
45. F. Molinet, J. Andronov, and D. Bouche, *Asymptotic and Hybrid Methods in Electromagnetics*, IEE (2005).
46. A. Skarlatos, R. Schuhmann, and T. Weiland, "Solution of Radiation and Scattering Problems in Complex Environments Using a Hybrid Finite Integration Technique-Uniform Theory of Diffraction Approach," *IEEE Trans. Ant. Propagat.*, vol. 53 (October 2005): 3347–3357.
47. A. Alu, F. Bilotti, and L. Vegni, "Method of Lines Analysis of Conformal Antennas," *IEEE Trans. Ant. Propagat.*, vol. 52 (June 2004): 1530–1540.

48. L. Josefsson and P. Persson, *Conformal Array Antenna Theory and Design*, IEEE (New York: Wiley, 2006).
49. D. H. Werner, R. J. Allard, R. A. Martin, and R. Mittra, "A Reciprocity Approach for Calculating Radiation Patterns of Arbitrarily-Shaped Microstrip Antennas Mounted on Circularly Cylindrical Platforms," *IEEE Trans. Ant. Propagat.*, vol. 51 (April 2003): 730–738.
50. A. Elsherbeni and M. Inman, *Antenna Design and Visualization using MATLAB* (Raleigh: NC: SciTech Publishing, 2006).
51. OpenMP web site: <http://www.openmp.org>.
52. Boston University, <http://scv.bu.edu/SCV/Tutorials/OpenMP/>.
53. Texas A&M web site, <http://sc.tamu.edu/help/>.
54. H. Jordan and G. Alaghband, *Fundamentals of Parallel Processing* (Upper Saddle River, NJ: Prentice Hall, 2003).
55. M.J. Quinn, *Parallel Programming in C with MPI and OpenMP* (New York: McGraw-Hill, 2004).
56. R. Chandra, R. Menon, L. Daqum, D. Kohr, D. Maydan, and J. McDonald, *Parallel Programming in OpenMP* (San Francisco: Morgan Kaufmann, 2000).
57. B. P. Lester, *The Art of Parallel Programming*, 2nd Ed. (Fairfield, IA: World Publishing, 2006).
58. P. Pacheco, *Parallel Programming in MPI* (San Francisco: Morgan Kaufmann, 1996).

